University of Zurich[UZH]

MASTER THESIS – Communication Systems Group, Prof. Dr. Burkhard Stiller

Create a Swiss Landscape Aesthetics Dataset for Decentralized Federated Learning

Yujie Tao
Zurich, Switzerland
Student ID: 23-744-378

Supervisor: Chao Feng, Andy Aidoo, Prof. Burkhard Stiller
Date of Submission: March 15, 2026

University of Zurich
Department of Informatics (IFI)
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland

Master Thesis
Communication Systems Group (CSG)
Department of Informatics (IFI)
University of Zurich
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland
URL: http://www.csg.uzh.ch/

Declaration of Independence

I hereby declare that I have composed this work independently and without the use of any aids other than those declared (including generative AI such as ChatGPT). I am aware that I take full responsibility for the scientific character of the submitted text myself, even if AI aids were used and declared (after written confirmation by the supervising professor). All passages taken verbatim or in sense from published or unpublished writings are identified as such. The work has not yet been submitted in the same or similar form or in excerpts as part of another examination.

Zürich, 3/15/2026

Yujie Tao

Signature of student

Abstract

Landscape aesthetics play an important role in tourism, environmental planning, and public perception of natural environments. Recent advances in computer vision have enabled automatic prediction of aesthetic scores from images, commonly referred to as Image Aesthetic Assessment (IAA). However, most existing approaches rely on centralized datasets and centralized model training, which may overlook regional variations in aesthetic perception and raise potential privacy concerns.

This thesis proposes a decentralized framework for predicting landscape aesthetic scores from geographically distributed image datasets. A Swiss landscape image dataset is constructed using publicly available sources and enriched with geographic metadata. Since the dataset lacks human-provided aesthetic ratings, a transfer learning approach based on a pre-trained convolutional neural network is employed to generate pseudo scenicness scores. The resulting dataset is then partitioned according to Swiss cantons to simulate decentralized data distributions.

To investigate collaborative model training under heterogeneous data conditions, a decentralized federated learning (DFL) architecture is implemented using the FLyra platform. In this framework, each canton acts as an independent training node that exchanges model parameters through peer-to-peer communication without relying on a central aggregation server.

Experimental evaluation compares centralized machine learning and decentralized learning approaches using standard evaluation metrics. The results provide insights into the feasibility of decentralized aesthetic prediction and highlight the potential of DFL for modeling region-dependent visual perception.

Abstrakt

Landschaftsästhetik spielt eine wichtige Rolle für Tourismus, Umweltplanung und die öffentliche Wahrnehmung natürlicher Umgebungen. Fortschritte im Bereich des Computer Vision ermöglichen inzwischen die automatische Vorhersage ästhetischer Bewertungen aus Bildern, ein Forschungsgebiet, das als Image Aesthetic Assessment (IAA) bezeichnet wird. Die meisten bestehenden Ansätze basieren jedoch auf zentralisierten Datensätzen und zentralisiertem Modelltraining. Dadurch können regionale Unterschiede in der ästhetischen Wahrnehmung nur eingeschränkt berücksichtigt werden und es entstehen potenzielle Datenschutzprobleme.

In dieser Arbeit wird ein dezentraler Ansatz zur Vorhersage von Landschaftsästhetik aus geografisch verteilten Bilddatensätzen vorgestellt. Hierzu wird zunächst ein Datensatz mit Schweizer Landschaftsbildern aus öffentlich verfügbaren Quellen erstellt und mit geografischen Metadaten angereichert. Da der Datensatz keine manuell vergebenen ästhetischen Bewertungen enthält, wird ein Transfer-Learning-Ansatz auf Basis eines vortrainierten Convolutional Neural Networks eingesetzt, um sogenannte pseudo-ästhetische Bewertungen zu erzeugen. Anschließend wird der Datensatz entsprechend der 26 Schweizer Kantone partitioniert, um eine dezentrale Datenverteilung zu simulieren.

Zur Untersuchung kollaborativer Trainingsverfahren unter heterogenen Datenbedingungen wird eine Architektur für Decentralized Federated Learning (DFL) auf der FLyra-Plattform implementiert. In diesem System fungiert jeder Kanton als unabhängiger Trainingsknoten, der Modellparameter über Peer-to-Peer-Kommunikation austauscht, ohne dass ein zentraler Aggregationsserver erforderlich ist.

Die experimentelle Evaluation vergleicht zentrale Machine-Learning-Ansätze mit dezentralen Lernverfahren anhand etablierter Evaluationsmetriken. Die Ergebnisse liefern Einblicke in die Machbarkeit dezentraler ästhetischer Vorhersagen und zeigen das Potenzial von DFL für die Modellierung regional unterschiedlicher visueller Wahrnehmungen.

Acknowledgments

I would like to express my sincere gratitude to my supervisors, Prof. Dr. Burkhard Stiller, Chao Feng, and Andy Aidoo, for their guidance, valuable feedback, and continuous support throughout the development of this thesis. Their insights and suggestions greatly contributed to shaping the direction and quality of this work.

I am also thankful to the members of the Communication Systems Group (CSG) at the University of Zurich for providing a stimulating research environment and helpful discussions during the course of this project.

Furthermore, I would like to thank my friends and fellow students for their encouragement and support during my studies.

Finally, I would like to express my deepest appreciation to my family for their unwavering support, patience, and encouragement throughout my academic journey.

Note: I have used ChatGPT (OpenAI) solely as a writing assistant for language refinement and clarity improvement. All scientific content, analysis, and conclusions are my own work.

Contents

Chapter 1

Introduction

1.1 Motivation

1.2 Motivation

In recent years, the rapid development of data-driven technologies has significantly increased the demand for large-scale machine learning systems capable of learning from distributed data sources. At the same time, growing concerns about data privacy and data ownership, together with regulations such as the General Data Protection Regulation (GDPR), have changed how machine learning systems are designed and deployed. Traditional centralized learning approaches typically require data from multiple sources to be collected and stored in a single location for model training. While this paradigm works well in controlled environments, it becomes increasingly problematic in distributed settings where privacy constraints, communication costs, and institutional boundaries prevent the free sharing of raw data [1].

Federated Learning (FL) has emerged as a promising solution to this challenge. Instead of transferring raw data to a central server, FL enables multiple clients to collaboratively train a shared model while keeping their data locally on their own devices or institutions. During training, each client performs local updates using its private data and periodically shares model parameters or gradients with a central server for aggregation [2]. This approach allows participants to benefit from collaborative learning while preserving data privacy.

However, most existing federated learning frameworks rely on a centralized aggregation server to coordinate model updates. Although this architecture reduces data sharing, it introduces several limitations. The central server may become a communication bottleneck as the number of clients increases, and it represents a single point of failure in the learning system. These limitations become particularly relevant in large-scale or geographically distributed environments.

To address these challenges, Decentralized Federated Learning (DFL) has been proposed as an alternative learning paradigm. In DFL systems, model updates are exchanged directly between participating nodes through peer-to-peer communication without relying on a central aggregation server [3], [4]. By distributing the aggregation process across the network, decentralized learning improves system robustness, scalability, and fault tolerance while preserving the privacy advantages of federated learning.

Despite the rapid progress in decentralized learning, most existing research has focused on objective machine learning tasks such as image classification, segmentation, or recommendation systems. Comparatively little attention has been given to subjective perception tasks, where labels are influenced by human preferences and cultural context. One representative example of such a task is Aesthetic Image Assessment (AIA), which aims to predict the perceived visual appeal of images.

Aesthetic evaluation differs from many traditional computer vision problems because it depends on subjective human judgment rather than objective ground truth. Large-scale datasets such as AVA and AADB have enabled the development of deep learning models capable of predicting aesthetic scores for images [5], [6]. However, these datasets mainly focus on general photography aesthetics and are typically used in centralized learning environments.

A dataset that focuses specifically on landscape aesthetics is the ScenicOrNot (SoN) dataset, which contains more than 200,000 geotagged landscape images across Great Britain with crowdsourced scenicness ratings [7]. While this dataset has enabled research on computational landscape perception, no comparable dataset currently exists for Switzerland, despite its diverse natural landscapes and strong regional variation.

The absence of such a dataset limits the ability to investigate regional differences in landscape perception and to study how decentralized learning systems behave under geographically heterogeneous data distributions. These challenges motivate the development of a new framework that combines region-aware landscape datasets with decentralized federated learning architectures.

1.3 Description of Work

This thesis investigates the application of decentralized learning to the problem of scenicness prediction using geographically distributed landscape image datasets. The goal is to design and implement a complete pipeline that enables decentralized model training on regionally partitioned image data while analyzing the behavior of the learning process under heterogeneous data distributions.

The overall workflow of this study consists of several stages, including dataset construction, pseudo label generation, decentralized training, and experimental analysis.

First, a supervised scenicness prediction model is trained using the ScenicOrNot dataset. ScenicOrNot is a crowdsourced dataset containing landscape photographs with human-provided scenicness ratings. These ratings represent the perceived aesthetic quality of the landscapes and serve as ground-truth labels for training a regression model. A deep learning model is trained to learn the relationship between visual features of landscape images and their corresponding scenicness scores.

After training, the model is used to generate scenicness predictions for a newly collected dataset of Swiss landscape images. These images are gathered from online repositories such as Wikimedia Commons and Flickr. Since the Swiss dataset does not contain human-annotated scenicness scores, the trained model is applied to produce predicted scores for each image. These predicted values serve as pseudo labels that enable further model training on the Swiss dataset.

To simulate a geographically distributed learning environment, the Swiss dataset is partitioned according to the 26 Swiss cantons. Each canton represents a decentralized training node containing a subset of the images associated with that region. Due to differences in image availability and geographic characteristics, the number of images varies across cantons, resulting in a naturally heterogeneous data distribution.

The decentralized training process is implemented using the FLyra framework. In this setting, each canton node performs local model updates using its regional dataset. During training, nodes exchange model parameters with neighboring nodes through peer-to-peer communication rather than relying on a centralized aggregation server. This decentralized setup allows the system to emulate realistic distributed learning conditions where data remain localized.

Finally, the training process and prediction results are analyzed to understand the behavior of decentralized learning under geographically distributed data. The evaluation focuses on several aspects, including training convergence behavior, scenicness score distributions, and regional variations in predicted aesthetic quality across Swiss cantons.

Through this pipeline, the thesis provides a practical implementation of decentralized learning for landscape aesthetic prediction and explores how geographic data heterogeneity influences model training and prediction outcomes.

1.4 Thesis Outline

The remainder of this thesis is organized as follows.

- **Chapter 2 – Background** introduces the fundamental concepts related to this work, including machine learning techniques for image analysis, federated learning, and decentralized federated learning. These concepts provide the theoretical foundation for the decentralized training framework used in this study.
- **Chapter 3 – Related Work** reviews existing research on image aesthetic assessment, landscape scenicness prediction, and applications of federated and decentralized learning in computer vision tasks.
- **Chapter 4 – Architecture / System Design** presents the overall architecture of the proposed system and describes the pipeline for dataset construction, pseudo-label generation, and decentralized learning.

- **Chapter 5 – Implementation** describes the practical implementation of the system, including dataset collection from Wikimedia Commons and Flickr, data preprocessing, scenicness prediction model training, and decentralized training using the FLyra platform.

- **Chapter 6 – Evaluation** presents the experimental results and analysis, including training convergence behavior, scenicness score distribution, and canton-level variations in predicted landscape aesthetic quality across Switzerland.

- **Chapter 7 – Summary and Conclusions** summarizes the main findings of this thesis, discusses the limitations of the current work, and outlines potential directions for future research.

Chapter 2

Background

This chapter introduces the fundamental concepts that form the theoretical foundation of this thesis. It first reviews the evolution from traditional centralized machine learning to federated learning and decentralized federated learning. The chapter then discusses optimization principles underlying decentralized training, followed by personalization strategies designed to address data heterogeneity in federated environments.

Finally, the chapter introduces the problem of aesthetic image assessment, which serves as the primary application domain for evaluating decentralized learning in this work.

To provide an overview of the learning paradigms discussed in this chapter, Figure 2.1 illustrates the evolution from centralized machine learning to federated learning and decentralized federated learning. In the traditional centralized learning paradigm, training data from multiple sources are collected and aggregated on a central server where a global model is trained. Federated learning extends this paradigm by allowing multiple clients to collaboratively train a shared model while keeping their data locally on each device or institution.

Decentralized federated learning further removes the need for a central aggregation server and enables peer-to-peer communication among participating nodes. Through iterative local training and model exchange, nodes collectively learn a global representation while preserving data locality and improving system robustness. In the context of this thesis, decentralized federated learning is applied to landscape scenicness prediction, where geographically distributed datasets from different Swiss cantons are used to collaboratively train aesthetic prediction models[8].

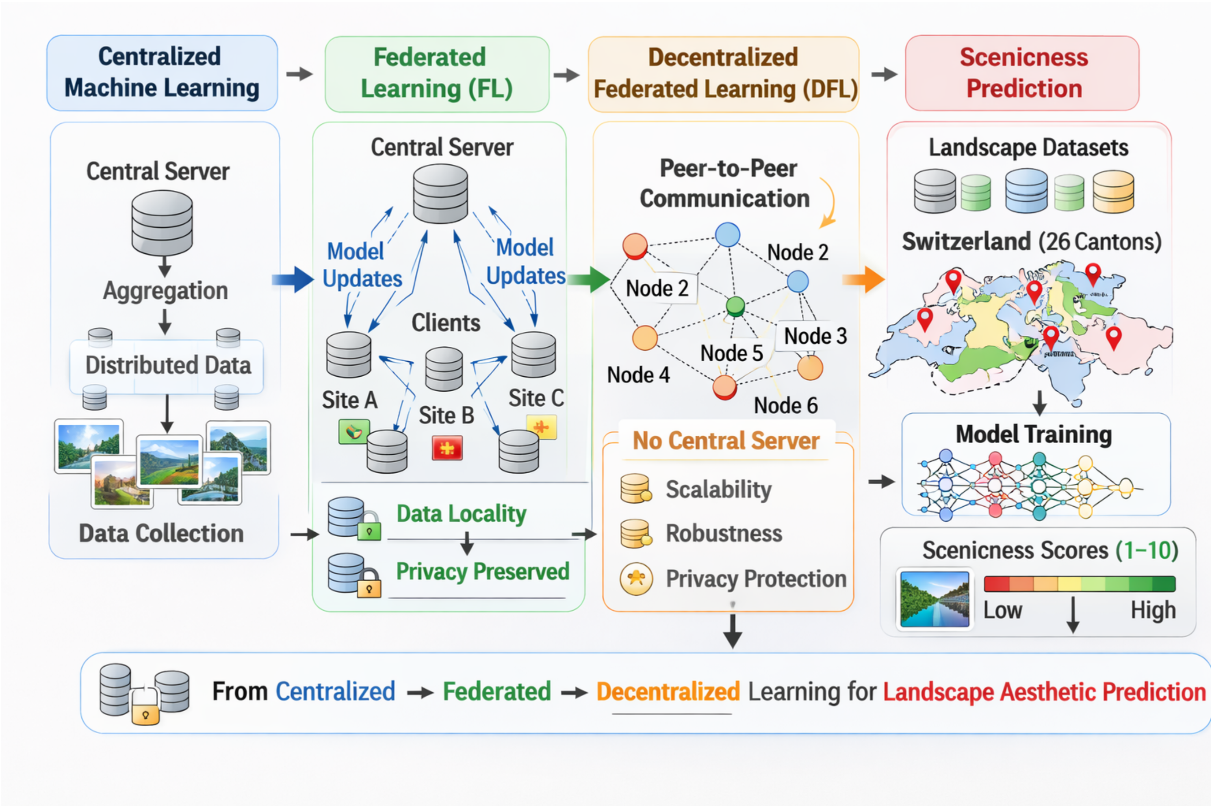

Figure 2.1: Conceptual overview of decentralized federated learning for landscape scenicness prediction.

2.1 From Centralized to Federated Learning

2.1.1 The Centralized Learning Paradigm

In the conventional paradigm of machine learning, training data from multiple sources are collected and stored in a single centralized repository. A global model is then optimized by minimizing a loss function $L(\theta)$ over the aggregated dataset. This approach assumes that data can be freely transmitted and stored without privacy or ownership concerns. While effective in data-rich environments such as large data centers, this paradigm has become increasingly impractical in the modern era of distributed data generation, where information is often confined to user devices, organizations, or sensors that cannot share raw data due to privacy regulations such as the General Data Protection Regulation (GDPR, 2018).

Furthermore, centralized training creates critical bottlenecks in communication and computation, as well as risks of data leakage and single-point failure. As edge devices proliferate, the amount of data produced outside data centers is expected to exceed 75% of the total global data volume by 2025 [9]. These developments have motivated the emergence of distributed and privacy-preserving training frameworks.

Although centralized learning has achieved remarkable success in many machine learning applications, it presents several limitations when applied to distributed data environments.

First, transferring large volumes of raw data to a central server introduces significant communication overhead. In scenarios where data are generated at the network edge, such as mobile devices or geographically distributed sensors, this process can become prohibitively expensive.

Second, centralized data aggregation raises serious privacy concerns. Sensitive data such as personal information, medical records, or location data cannot always be shared due to regulatory constraints and institutional policies. Consequently, many organizations are unable or unwilling to contribute their data to centralized machine learning pipelines.

Finally, centralized training architectures introduce a single point of failure. If the central server becomes unavailable or compromised, the entire learning process may be disrupted. These challenges have motivated the development of distributed and privacy-preserving learning frameworks such as federated learning.

2.1.2 Fundamentals of Federated Learning (FedAvg)

Federated Learning (FL) was introduced to enable collaborative model training without requiring participants to share their raw data. In their seminal work *Communication-Efficient Learning of Deep Networks from Decentralized Data*, McMahan *et al.* [2] proposed the **Federated Averaging (FedAvg)** algorithm, which remains the foundation of most FL systems.

The FedAvg procedure operates iteratively in three steps:

1. The server initializes and broadcasts a global model θ_t to a selected subset of clients.
2. Each client performs several local updates on its private data D_k using stochastic gradient descent (SGD):

$$\theta_{t+1}^{(k)} \leftarrow \theta_t - \eta \nabla L_k(\theta_t; D_k)$$

3. The clients send their updated parameters to the server, which computes a weighted average:

$$\theta_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n_{\text{total}}} \theta_{t+1}^{(k)}$$

   where n_k is the local sample size for client k.

This simple yet effective averaging scheme allows the global model to incorporate knowledge from diverse local datasets while preserving data locality. Subsequent studies **kairouz2021advance** [10] demonstrated that FedAvg can achieve near-centralized performance under ideal conditions while drastically reducing communication costs.

2.1.3 Challenges in Federated Learning

Despite its promise, federated learning introduces several unique challenges that differentiate it from classical distributed learning **kairouz2021advances**.

(a) Statistical heterogeneity. Client datasets are typically non-independent and non-identically distributed (non-IID). This results in client drift, where local updates diverge from the global optimum, leading to slower or unstable convergence. Li et al. [10] addressed this issue through FedProx, which adds a proximal term to stabilize local training.

(b) System heterogeneity. Clients vary in computational resources, battery life, and connectivity, causing stragglers or dropouts during communication rounds. Techniques such as adaptive client selection [11] mitigate these disparities.

(c) Communication constraints. Bandwidth limitations make frequent model exchanges costly. Compression and quantization methods, such as those surveyed by Sattler et al. [12], reduce the size of transmitted updates.

(d) Privacy and security. Although data remain local, gradient updates may leak sensitive information. Differential privacy [13] and secure aggregation protocols [14] are widely adopted countermeasures.

These challenges highlight that federated learning is not merely distributed optimization but a fundamentally new paradigm balancing three conflicting objectives: model accuracy, communication efficiency, and privacy protection. The next section extends this discussion toward **decentralized federated learning (DFL)**, which removes the central server to further enhance scalability and fault tolerance [15].

In practice, these challenges often interact with each other. For example, statistical heterogeneity can amplify the impact of system heterogeneity, leading to unstable convergence in large-scale deployments. Similarly, communication constraints may limit the frequency of model aggregation, which further exacerbates divergence between local and global models. Therefore, designing scalable and robust federated learning systems requires careful consideration of both algorithmic and system-level factors.

Table 2.1: Comparison of centralized machine learning, federated learning, and decentralized federated learning.

Feature	**Centralized ML**	**Federated Learning**	**Decentralized FL**
Data Location	Central server	Local clients	Local nodes
Model Aggregation	Central training	Server aggregation	Peer-to-peer aggregation
Privacy Protection	Low	Medium	High
Scalability	Limited	Moderate	High
Single Point of Failure	Yes	Yes	No
Communication Pattern	Data transfer	Client-server	Peer-to-peer
Typical Application	Traditional ML	Mobile / Edge FL	Distributed systems

To summarize the key differences between centralized machine learning, federated learning, and decentralized federated learning, Table 2.1 provides a conceptual comparison of their main characteristics.

2.2 Decentralized Federated Learning (DFL)

2.2.1 Motivation for Decentralization

While the classical federated learning architecture relies on a central server to aggregate model updates, this design still introduces a potential bottleneck and single point of failure. In large-scale or cross-institutional collaborations, the central coordinator may be unavailable or untrusted, and the communication cost for all clients to interact with a single node becomes prohibitive. To overcome these limitations, **decentralized federated learning (DFL)** removes the central server and allows clients to exchange information directly with their neighbors [3], [4].

Decentralization not only improves scalability and robustness but also enhances privacy preservation, since no central node has access to the complete model updates. In addition, DFL is more aligned with realistic networking environments such as peer-to-peer systems, edge networks, and blockchain-based infrastructures, where nodes form dynamic and partially connected topologies [16], [17].

2.2.2 Network Topologies and Gossip Mechanisms

In DFL, participating clients are connected by a communication graph $G = (V, E)$, where V denotes the set of clients and E the set of communication links. Each node $i \in V$ maintains its own local model θ_i and periodically averages it with models received from its neighbors N_i. The update rule can be expressed as:

$$\theta_i^{t+1} = \sum_{j \in N_i \cup \{i\}} w_{ij}\, \theta_j^t - \eta \nabla L_i(\theta_i^t; D_i)$$

where w_{ij} represents the weight of the connection between nodes i and j, satisfying $\sum_j w_{ij} = 1$. Such communication protocols are often referred to as *gossip learning* [18], [19].

Different topologies (e.g., ring, mesh, random, fully-connected) lead to varying trade-offs between convergence speed and communication efficiency. Empirical studies show that well-designed sparse graphs can achieve performance close to centralized federated learning while using far fewer communication links [4], [20]. Recent frameworks such as D-PSGD [3] and AD-PSGD [21] demonstrate that asynchronous gossip averaging further improves robustness to network latency and client dropouts.

2.2.3 Convergence and Robustness in DFL

From a theoretical standpoint, DFL algorithms aim to solve the consensus optimization problem:

$$\min_{\theta_1,\ldots,\theta_K} \frac{1}{K} \sum_{k=1}^{K} f_k(\theta_k) \quad \text{s.t.} \quad \theta_i = \theta_j, \forall (i,j) \in E$$

Under standard assumptions of smoothness and bounded variance, D-PSGD [3] achieves an $\mathcal{O}(1/\sqrt{KT})$ convergence rate comparable to centralized SGD. Koloskova *et al.* [20] proved that even with overlapping consensus steps, decentralized training can converge to stationary points with rates close to fully synchronous methods.

Regarding robustness, the absence of a central aggregator eliminates the single-point failure issue inherent in traditional FL. However, decentralized schemes may suffer from slower convergence on poorly connected graphs or when malicious clients inject false updates. Research on Byzantine-resilient DFL [17], [22] addresses these challenges through robust aggregation and adaptive weighting strategies.

In summary, DFL offers a promising paradigm that balances scalability, robustness, and privacy by replacing the central aggregation step with distributed peer-to-peer communication. The following section introduces the optimization principles underlying such decentralized algorithms.

2.3 Optimization Principles in Decentralized Learning

2.3.1 Consensus Optimization Formulation

Decentralized federated learning can be interpreted as a distributed optimization problem where each client k possesses a local objective function $f_k(\theta)$ defined on its private dataset D_k. The global learning goal is to minimize the average empirical risk across all clients:

$$\min_{\theta} F(\theta) = \frac{1}{K} \sum_{k=1}^{K} f_k(\theta)$$

Unlike centralized optimization, in decentralized settings each node maintains its own local parameter θ_k and exchanges information only with its neighbors. The corresponding optimization problem can be expressed as a **consensus optimization** [23]:

$$\min_{\theta_1,\ldots,\theta_K} \frac{1}{K} \sum_{k=1}^{K} f_k(\theta_k) \quad \text{s.t.} \quad \theta_i = \theta_j, \, \forall (i,j) \in E$$

where E denotes the set of communication links in the network graph. The equality constraints ensure that all local models eventually reach a consensus. Such a formulation enables gradient-based methods to operate in a decentralized manner while asymptotically achieving global optimality under mild assumptions.

2.3.2 Local Gradient Descent and Averaging

To solve the above consensus problem, early works such as Distributed Gradient Descent (DGD) [23] and its stochastic variants [3], [24] proposed a two-step update rule: each node performs a local gradient descent step followed by a model averaging step with its neighbors:

$$\theta_i^{t+1} = \sum_{j \in N_i \cup \{i\}} w_{ij}\, \theta_j^t - \eta_t \nabla f_i(\theta_i^t)$$

where w_{ij} is the (i, j) entry of a symmetric, doubly stochastic weight matrix W that defines the mixing topology. The matrix W governs the information propagation speed: a denser connectivity or higher spectral gap leads to faster consensus. Lian *et al.* [3] established that this decentralized parallel stochastic gradient descent (D-PSGD) algorithm achieves an $\mathcal{O}(1/\sqrt{KT})$ convergence rate comparable to centralized SGD. Subsequent research extended this framework to handle asynchronous communication (AD-PSGD) [21], time-varying graphs [25], and compression [26] to further reduce communication overhead.

2.3.3 Convergence Rates and Efficiency

Theoretical analysis of decentralized optimization typically relies on the *spectral properties* of the mixing matrix W. When W is symmetric and doubly stochastic with spectral radius $\rho(W - \frac{1}{K}\mathbf{1}\mathbf{1}^T) < 1$, the local iterates θ_i^t are guaranteed to converge to the same stationary point [20], [24]. In practice, the convergence speed depends jointly on the learning rate η_t, network connectivity, and gradient variance across clients. Koloskova *et al.* [20] showed that overlapping consensus steps allow DFL to achieve communication efficiency close to centralized FedAvg even under sparse topologies.

Recent work has also focused on improving scalability and robustness of decentralized optimization. Gradient tracking methods such as Exact Diffusion [27] and D2 [26] correct the bias introduced by local averaging and significantly accelerate convergence. Moreover, asynchronous variants [21] eliminate synchronization barriers, allowing high-latency or heterogeneous clients to participate without degrading overall performance. These advances have established decentralized optimization as a mature theoretical foundation for scalable and fault-tolerant federated learning systems [28].

The convergence behavior of decentralized optimization algorithms depends strongly on the structure of the communication graph connecting participating nodes. Well-connected graphs allow information to propagate quickly across the network, enabling faster consensus between local models. In contrast, sparse communication graphs may slow down convergence but reduce the overall communication cost [29].

In practical decentralized learning systems, the design of communication topology therefore represents an important trade-off between efficiency and scalability [30]. Many studies have shown that carefully designed sparse graphs can achieve performance close to centralized training while requiring significantly fewer communication rounds [31].

2.4 Personalization in Federated Learning

2.4.1 Motivation for Personalized FL

A core assumption of standard federated learning is that a single global model can generalize well across all clients. However, in practice, client data distributions are often *non-independent and non-identically distributed* (non-IID) due to differences in user behavior, environment, or sensor domain. Such heterogeneity causes the global model to overfit to dominant data sources and underperform on minority or outlier clients [1], [10]. This phenomenon, commonly referred to as *client drift*, leads to unstable convergence and degraded personalization.

To address this issue, personalized federated learning (pFL) aims to strike a balance between the advantages of global collaboration and the specificity of local adaptation. Instead of enforcing complete consensus among clients, pFL allows each participant to maintain a customized model while sharing certain representations or gradients with others [32]. This approach is especially relevant for subjective tasks such as aesthetic evaluation, where the target distribution inherently varies across users or regions.

2.4.2 Representative Algorithms

A number of algorithms have been proposed to incorporate personalization into the FL framework. Among the most widely studied are:

- **FedProx** [10]: Introduces a proximal term to the local objective function $f_k(\theta) + \frac{\mu}{2}\|\theta - \theta_t\|^2$, which constrains local updates and mitigates divergence caused by non-IID data. It improves training stability in heterogeneous systems and serves as a practical baseline for many pFL methods.

- **Per-FedAvg** [33]: Inspired by model-agnostic meta-learning (MAML), Per-FedAvg learns an initialization that can be quickly adapted to each client using a few local gradient steps. This method yields strong personalization with minimal additional computation.

- **pFedMe** [34]: Formulates personalization as a bi-level optimization problem based on Moreau envelope regularization, allowing each client to maintain its own personalized parameter θ_k close to a shared global model. This approach provides theoretical convergence guarantees under non-IID data.

- **FedPer and FedRep** [35], [36]: Decompose the model into a shared feature extractor and client-specific classifier heads, enabling representation sharing while allowing flexible personalization at the output layer. These methods perform well on heterogeneous visual tasks.

2.4.3 Trade-offs Between Global and Local Models

Personalized FL introduces an inherent trade-off between global generalization and local specialization. A fully shared global model ensures collaboration but may fail to capture client-specific patterns, whereas entirely independent models sacrifice knowledge transfer. The effectiveness of pFL algorithms depends on how well they balance these competing bjectives [33], [36].

From an optimization perspective, the trade-off can be viewed as minimizing a combination of global loss and local deviation:

$$\min_{\{\theta_k\}} \sum_{k=1}^{K} \left[f_k(\theta_k) + \lambda \|\theta_k - \bar{\theta}\|^2 \right],$$

where $\bar{\theta}$ denotes the shared global representation and λ controls the degree of personalization. This formulation unifies several existing approaches including FedProx and pFedMe.

For subjective domains such as aesthetic image assessment, personalization is crucial because user preferences and visual perceptions vary across individuals and regions. A personalized DFL framework enables each node (e.g., Swiss canton) to adapt the shared aesthetic model to its local visual characteristics while benefiting from knowledge transfer across the entire network.

Personalization strategies are particularly important in applications where data distributions vary significantly across clients. In such scenarios, enforcing a single global model may lead to suboptimal performance for many participants.

By allowing local adaptation while preserving shared knowledge across clients, personalized federated learning methods provide a flexible framework for handling heterogeneous datasets. This property is especially relevant for subjective perception tasks such as aesthetic evaluation, where individual preferences and regional characteristics may differ substantially.

2.5 Aesthetic Image Assessment (AIA)

2.5.1 Problem Definition and Relevance

Aesthetic Image Assessment (AIA) aims to automatically predict the perceived visual appeal or aesthetic quality of an image. The task is inherently subjective, as it depends on human perception, composition, color harmony, and semantic content. With the rapid growth of image-sharing platforms, AIA has become an important problem in computer vision, enabling applications such as photo recommendation, automated curation, and visual enhancement [6], [37].

Formally, AIA can be formulated as a supervised regression or classification task. Given an image I_i and its associated human rating $y_i \in [1, 10]$, a model f_θ is trained to approximate the mapping $f_\theta(I_i) \approx y_i$. The challenge lies in learning robust representations that generalize across diverse content types and subjective preferences. In this thesis, AIA serves as the application domain for evaluating decentralized federated learning frameworks, highlighting how DFL can handle data heterogeneity and subjective annotation bias.

2.5.2 Major Datasets and Benchmarks

Early research on aesthetic modeling was enabled by large-scale public datasets that collected human aesthetic ratings from photography communities and online crowdsourcing. Several datasets have been developed to support research in aesthetic image assessment. These datasets differ in terms of image content, rating methodology, and intended research applications. Table 2.2 summarizes several representative datasets commonly used in aesthetic prediction studies.

- **AVA (Aesthetic Visual Analysis)** [5]: Introduced by Murray *et al.* at CVPR 2012, AVA is the largest and most influential benchmark in AIA, containing over 255,000 images from the DPChallenge online community. Each image is annotated with 200–250 aesthetic ratings on a 1–10 scale, along with photographic style and semantic labels. The dataset established the standard regression and classification setups for aesthetic score prediction and is still widely used today.

- **AADB (Aesthetics and Attributes Database)** [6]: Collected from Flickr and photo-sharing websites, AADB contains approximately 10,000 images annotated with both aesthetic scores and 11 interpretable attributes (e.g., color harmony, lighting, composition). It emphasizes attribute-level reasoning, enabling fine-grained learning of visual aesthetics.

- **NIMA (Neural Image Assessment)** [38]: Proposed by Google Research, NIMA is not a dataset but a deep CNN model trained on AVA and other sources to predict score distributions rather than mean ratings. It demonstrates that predicting the full rating histogram yields more stable and human-aligned results, inspiring subsequent probabilistic models.

- **KonIQ-10k** [37]: Originally designed for image quality assessment, KonIQ-10k includes 10,073 images with subjective ratings collected via crowdsourcing on real-world photos. Due to its visual diversity and realistic distortions, it is also frequently used for aesthetic evaluation tasks.

- **ScenicOrNot (SoN)** [7]: A unique dataset of over 200,000 geotagged landscape images across Great Britain with crowdsourced "scenicness" ratings from 1 to 10, collected via the website ScenicOrNot.com. It has been used in environmental psychology and computational geography to study the relationship between scenic beauty, natural environments, and human well-being. Its geospatial diversity makes it an ideal reference for constructing region-specific aesthetic datasets, such as the Swiss landscape dataset proposed in this work.

Table 2.2: Comparison of commonly used datasets for aesthetic image assessment

Dataset	Images	Image Type	Rating Type	Typical Usage
AVA	∼250,000	Photography	Aesthetic score distribution	Deep learning for aesthetic prediction
AADB	10,000	Photography	Aesthetic scores + attributes	Attribute-aware aesthetic modeling
KonIQ-10k	10,073	Real-world images	Quality and aesthetic ratings	Image quality and aesthetic evaluation
ScenicOrNot	∼200,000	Landscape images	Scenicness scores (1–10)	Landscape perception and environmental analysis

These datasets collectively form the foundation for modern AIA research. While AVA and AADB focus on artistic photography, SoN emphasizes geographic and environmental context, aligning closely with the objective of modeling regional landscape aesthetics under privacy-aware and decentralized learning conditions. Among these datasets, ScenicOrNot is particularly relevant for landscape aesthetics research because it contains geotagged outdoor images with crowdsourced scenicness ratings. However, it focuses exclusively on landscapes in Great Britain, which motivates the development of a dedicated Swiss landscape aesthetics dataset in this thesis.

2.5.3 Learning Methods for Aesthetic Prediction

Traditional aesthetic prediction relied on handcrafted visual features such as color histograms, rule-of-thirds composition, and saliency maps [39]. With the advent of deep learning, convolutional neural networks (CNNs) have become the dominant approach for learning high-level aesthetic representations directly from images. The AVA and AADB datasets enabled end-to-end CNN training for aesthetic score regression and ranking tasks.

The NIMA model [38] introduced the idea of predicting a distribution over aesthetic scores using a softmax activation and Earth Mover's Distance (EMD) loss, rather than predicting a single scalar rating. More recent approaches leverage multi-task and attribute-aware frameworks [6], graph-based representations [37], and contrastive learning for personalized aesthetic preference modeling.

Despite remarkable progress, existing datasets are primarily centralized and may reflect demographic or cultural biases from specific populations. In contrast, a decentralized learning setup allows aesthetic models to be collaboratively trained across different regions or institutions without sharing raw images, enabling fairer and more representative modeling of diverse visual preferences. This thesis therefore proposes building a Swiss Landscape Aesthetic Dataset and integrating it into a decentralized federated learning framework to evaluate model performance and robustness under realistic privacy constraints [40].

Despite the significant progress achieved by deep learning approaches, aesthetic prediction remains a challenging problem. Unlike objective computer vision tasks, aesthetic evaluation is inherently subjective and influenced by individual perception, cultural background, and contextual factors‘[41].

Furthermore, most existing aesthetic datasets are collected from specific online communities or geographic regions. As a result, models trained on these datasets may capture only limited aspects of global aesthetic preferences‘[42]. This limitation motivates the exploration of decentralized learning approaches that allow aesthetic models to be trained across geographically distributed datasets while preserving data locality [43].

In summary, decentralized federated learning provides a promising paradigm for collaborative model training in distributed environments where data cannot be centralized. At the same time, aesthetic image assessment represents a challenging application domain characterized by subjective perception and heterogeneous data sources.

The combination of these two research directions creates an interesting opportunity to investigate how decentralized learning systems behave when applied to geographically distributed aesthetic datasets.

Chapter 3

Related Work

Federated learning has recently been applied to various computer vision and remote sensing tasks where training data are naturally distributed across multiple organizations or geographic regions. These applications include scene classification, semantic segmentation, and environmental monitoring using satellite imagery.

Existing studies primarily focus on objective vision tasks and rely on centralized federated learning architectures. The following section reviews representative approaches and highlights their limitations in handling geographically heterogeneous datasets.

3.1 Federated Learning in Vision and Remote Sensing Applications

Federated learning (FL) has rapidly evolved beyond its initial applications in mobile and healthcare domains to encompass various computer vision and remote sensing tasks [1]. The paradigm is particularly appealing for earth observation and landscape analysis, where data are geographically distributed across institutions, satellites, or national agencies that cannot freely exchange raw imagery due to privacy, bandwidth, or legal constraints. This section reviews representative FL approaches developed for vision and remote-sensing applications, with a focus on their objectives, methods, and limitations relevant to the present study.

3.1.1 Federated Scene Classification

Ben Youssef *et al. al.* [44] introduced a federated framework for remote-sensing scene classification using convolutional and transformer-based architectures. The authors employed three benchmark datasets and simulated multiple clients to evaluate communication efficiency and robustness to client dropouts. Their experiments demonstrated that Vision Transformers (ViT-Base) consistently outperform CNNs such as EfficientNet-B1/B3 under FL settings, particularly when some clients have incomplete participation. This work confirmed that FL can achieve competitive accuracy in geographically partitioned image data without centralizing raw imagery. However, its focus remained on objective classification tasks rather than subjective or perceptual evaluations.

3.1.2 Geographic Heterogeneity in FL Segmentation

Tan *et al.* [45] highlighted that traditional FL algorithms often degrade when applied to geographically heterogeneous remote-sensing data. They proposed the **GeoFed** framework, a geographic heterogeneity-aware federated learning method for semantic segmentation. GeoFed introduces three modules: (1) *Global Insight Enhancement* (GIE) to correct class-distribution imbalance, (2) *Essential Feature Mining* (EFM) to mitigate object appearance diversity across regions, and (3) *Local–Global Balance* (LoGo) to adjust personalization strength between clients. Experiments on three federated remote-sensing benchmarks (FedFBP, FedCASID, and FedInria) showed that GeoFed achieves superior cross-regional generalization compared with FedAvg and FedProx. The study underscores the necessity of incorporating geographic context into federated vision models—an insight particularly relevant for landscape aesthetic tasks, where regional perception differences are intrinsic.

3.1.3 Surveys on FL for Remote Sensing

Moreno-Álvarez *et al.* [46] provided a comprehensive survey summarizing how FL has been adopted for remote-sensing and earth-observation applications, including land-cover classification, crop mapping, and environmental monitoring. The authors categorize challenges into three main aspects: *data heterogeneity*, *communication efficiency*, and *privacy compliance*. Their analysis reveals that most FL studies in remote sensing rely on centralized server aggregation (FedAvg) and focus on objective tasks. They call for future work to address non-IID geographic distributions, limited communication infrastructure in remote areas, and privacy-preserving collaboration between agencies. This perspective motivates the present research to extend FL from objective scene understanding to subjective aesthetic modeling in a geographically decentralized context.

3.2 Decentralized and Personalized Federated Learning

While conventional federated learning (FL) frameworks rely on a central server to aggregate model updates from distributed clients, this architecture introduces potential bottlenecks related to communication overhead, server trust, and single-point failures. To address these limitations, decentralized and personalized variants of FL have been developed to improve scalability, robustness, and adaptability to non-IID (non-independent and identically distributed) data. This section reviews representative advances in decentralized FL (DFL) and personalized FL (pFL), summarizing their principles, mechanisms, and implications for subjective and region-aware learning tasks.

3.2.1 Decentralized Federated Learning (DFL)

In contrast to traditional server-based FL, decentralized federated learning (DFL) removes the need for a central coordinator by enabling clients to communicate and exchange model parameters directly with their peers. Early work by Lian *et al.* [47] introduced the *Decentralized Parallel Stochastic Gradient Descent (D-PSGD)* algorithm, which allows each node to update its model using locally available data and aggregate information from neighboring nodes through a communication graph. The approach proved that convergence can still be guaranteed under synchronous or asynchronous settings, even in the absence of a global server.

Subsequent extensions have incorporated more complex topologies and communication schemes. For instance, He extitet al. [48] proposed a *Gossip-based DFL* method that relies on random peer-to-peer message exchanges, reducing dependency on fully connected networks and improving scalability. Other works such as Koloskova *et al.* [49] and Tang *et al.* [26] provided unified theoretical frameworks for analyzing convergence and robustness under different network topologies. These DFL approaches demonstrate that communication efficiency and resilience can be achieved without centralized orchestration, making them particularly relevant for geographically distributed visual or sensor data where direct peer collaboration is feasible.

3.2.2 Personalized Federated Learning (pFL)

Although standard FL assumes all clients share the same global objective, in practice, local data distributions often vary significantly due to user behavior, domain bias, or regional heterogeneity. This non-IID challenge motivates personalized federated learning (pFL), which aims to tailor models to each client's specific data while still leveraging shared global knowledge.

One of the earliest methods, textitFedProx [10], introduced a proximal regularization term to stabilize local updates and mitigate client drift. Later, *Per-FedAvg* [33] adopted a meta-learning perspective, treating global aggregation as a meta-optimization process that helps clients quickly adapt to their own data. Alternative approaches such as textitpFedMe [50] and *FedRep* [36] further decouple shared representations from client-specific parameters, allowing local fine-tuning while preserving a compact shared model.

These methods significantly improve model performance in eterogeneous environments and have been widely applied to domains such as healthcare, speech, and user-behavior modeling. However, most of them have not yet been explored in subjective or perceptual domains, where heterogeneity originates not only from data distribution but also from individual or cultural preferences.

3.2.3 Implications for This Thesis

Overall, decentralized and personalized federated learning methods aim to address two major challenges of federated training: system scalability and data heterogeneity. Decentralized learning removes the reliance on a central aggregation server, while personalized learning allows models to adapt to client-specific data distributions.

However, most existing studies focus on objective prediction tasks and have not yet been extensively applied to subjective perception problems such as aesthetic image assessment.

Both DFL and pFL provide critical insights for the current research. DFL aligns with the goal of removing centralized aggregation, enabling peer-to-peer model sharing across geographically distributed clients, such as different Swiss regions in this thesis. Meanwhile, pFL emphasizes personalization under non-IID settings, which parallels aesthetic variation across linguistic and cultural boundaries. Combining these paradigms, the proposed study investigates Decentralized Federated Learning for regional aesthetic perception, extending the principles of robustness and personalization from objective tasks to subjective, human-centered modeling.

3.3 Aesthetic Image Assessment and Subjective Perception

Image Aesthetics Assessment (IAA) aims to quantify the perceived beauty or visual quality of images based on human judgments. Unlike objective computer vision tasks such as classification or segmentation, aesthetic evaluation is inherently subjective and is influenced by individual preferences, cultural backgrounds, and contextual factors. As a result, aesthetic perception often exhibits large variability across individuals and populations.

Early research in aesthetic assessment relied on handcrafted features derived from photographic rules, such as color harmony, composition balance, and depth cues. With the rapid development of deep learning, recent approaches have shifted toward data-driven models trained on large-scale aesthetic datasets [38]. These methods enable neural networks to learn complex visual patterns associated with aesthetic perception directly from image data.

Large annotated datasets have played a key role in advancing aesthetic prediction. One of the earliest and most influential datasets is the AVA (Aesthetic Visual Analysis) dataset, introduced by Murray et al. [5], which contains more than 250,000 images collected from the DPChallenge photography platform and annotated with aesthetic ratings from multiple users. Such datasets allow the development of machine learning models capable of predicting aesthetic scores from visual content.

More recent work has explored deep neural architectures and multi-modal models for aesthetic prediction. For instance, the Neural Image Assessment (NIMA) model predicts the full distribution of aesthetic ratings rather than a single mean score, providing a more nuanced representation of subjective aesthetic perception [38]. In addition, modern approaches increasingly incorporate multi-modal information such as textual descriptions or semantic attributes extracted from vision-language models like CLIP [51]. These developments highlight the growing complexity of aesthetic modeling and its reliance on large-scale datasets and deep learning frameworks.

This section reviews representative developments in aesthetic image assessment and discusses their relevance to decentralized federated learning settings.

3.3.1 Traditional Aesthetic Assessment Datasets and Models

The earliest benchmarks for computational aesthetic analysis were built using large-scale datasets annotated with human ratings. Among them, the AVA dataset [5] remains one of the most widely used resources for training aesthetic prediction models. The dataset contains over 250,000 images with aesthetic ratings collected from photography enthusiasts, allowing researchers to train models that learn correlations between visual features and perceived beauty.

Subsequent datasets further expanded the scope of aesthetic analysis. For example, the AADB dataset [6] introduced higher-quality aesthetic annotations collected under controlled evaluation protocols, while KonIQ-10k [37] provided a large-scale dataset focusing on image quality and perceptual attributes. These datasets enriched the diversity of aesthetic annotations and enabled more robust model training.

Early deep learning models for aesthetic assessment employed convolutional neural networks (CNNs) to extract visual features from images. The NIMA model [38] represents a notable example, using an Inception-v2 architecture to predict the probability distribution of aesthetic ratings rather than a single scalar score. This approach demonstrated that deep features can effectively capture subjective aesthetic dimensions.

Beyond predicting a single aesthetic score, several studies have explored ranking-based approaches for modeling aesthetic preferences. Instead of estimating absolute aesthetic scores, ranking models learn pairwise comparisons between images and determine which image is visually more appealing. Lu et al. [52] proposed a deep multi-patch aggregation network that captures local aesthetic features and learns relative aesthetic ranking between images. Such approaches are particularly suitable for subjective perception tasks because human judgments are often expressed through relative comparisons rather than absolute numerical scores. These developments demonstrate that aesthetic prediction models can capture complex visual patterns related to human perception when trained on large annotated datasets.

Despite these advances, most existing aesthetic datasets and models are trained in centralized settings. Consequently, they often reflect global averages of aesthetic preference and do not account for regional or cultural differences in perception.

3.3.2 Cross-Cultural and Regional Aesthetic Studies

Beyond individual taste, aesthetic judgment is also shaped by geographic, cultural, and socio-economic factors. Previous studies have demonstrated that perceptions of landscape beauty vary significantly across regions and populations.

For instance, Seresinhe et al. [53] introduced the ScenicOrNot dataset, which links geotagged landscape photographs from the United Kingdom to crowdsourced scenicness ratings. Their analysis revealed systematic differences between urban and rural scenery preferences and highlighted the importance of geographic context in aesthetic perception.

Related projects such as PlacePulse [54] investigated how urban environments are perceived in terms of safety, wealth, and beauty using crowdsourced visual comparisons. These studies demonstrated that aesthetic perception is closely related to spatial and socio-economic characteristics.

Such findings suggest that aesthetic perception exhibits geographic heterogeneity, analogous to the non-IID data distributions commonly encountered in federated learning systems. Consequently, models trained on globally aggregated datasets may fail to capture regional aesthetic preferences, motivating the development of region-aware learning frameworks.

3.3.3 Multi-Modal and Personalized Aesthetic Modeling

Recent research in IAA has increasingly focused on personalization and multi-modal modeling. These approaches aim to capture user-specific aesthetic preferences and incorporate multiple sources of information beyond raw image pixels.

Yang et al. [55] proposed the PAPA/PARA framework (Personalized Aesthetic Ranking with Attributes), which integrates user attributes and image features to predict personalized aesthetic rankings. By modeling user-specific preferences, the framework demonstrates improved performance compared with global aesthetic predictors.

More recently, multi-modal architectures have been introduced to combine visual and textual representations. For example, the Multi-modal Learnable Queries (MMLQ) model integrates visual features from CLIP [51] with textual representations from language models such as BERT. Through cross-attention mechanisms, these models capture complex relationships between visual content and semantic descriptions, achieving state-of-the-art performance on aesthetic prediction tasks.

However, most of these models remain centralized and require access to all user data during training. This requirement raises potential privacy concerns and limits scalability in distributed environments.

3.3.4 Federated Learning for Personalized Aesthetic Assessment

To address privacy concerns in aesthetic learning, Xiong *et al.* [56] proposed **FedPIAA**, the first federated framework for personalized image aesthetics assessment. FedPIAA enables collaborative model training across user devices without sharing raw images. Each client fine-tunes a lightweight CNN-based predictor on local data while the global model is aggregated using the FedAvg algorithm. Two configurations were proposed: (1) *FedPIAA1*, where all users participate in federated updates; and (2) *FedPIAA2*, where unseen users adapt locally after global convergence. Experiments on the Flickr-AES dataset ($\approx$ 41k images from 210 users) showed that FedPIAA improved the Spearman rank correlation (SRCC) by 1.5–4.8% over standard FedAvg, achieving accuracy comparable to centralized personalized models while preserving data privacy. However, the method remains centralized in orchestration and models *individual* user preferences rather than *regional* aesthetic variation.

Recent research has demonstrated that aesthetic perception is influenced by multiple factors including cultural background, geographic environment, and individual preferences. These findings suggest that aesthetic prediction models trained on centralized datasets may not generalize well across different regions.

Consequently, decentralized learning frameworks provide an attractive opportunity for modeling aesthetic perception across geographically distributed datasets while preserving data locality.

3.3.5 Limitations and Motivation for This Thesis

Another important challenge in aesthetic modeling lies in the heterogeneous nature of aesthetic preferences. Unlike objective visual recognition tasks, aesthetic perception can vary significantly across individuals and geographic regions. Factors such as cultural background, environmental characteristics, and personal experiences may influence how people perceive the beauty of landscapes. Consequently, models trained on globally aggregated datasets may fail to capture regional variations in aesthetic perception. This limitation highlights the potential benefits of distributed learning approaches that allow models to adapt to heterogeneous data distributions across different locations.

Although existing research has significantly advanced automated aesthetic assessment, several limitations remain.

First, nearly all aesthetic datasets and models operate under centralized paradigms, requiring global data aggregation and limiting privacy protection. Second, current aesthetic learning approaches mainly focus on global averages or individual preferences, leaving geographic and cultural heterogeneity largely unexplored. Finally, only limited work has investigated federated learning for aesthetic prediction, and none has examined decentralized federated learning (DFL) in this context.

This thesis addresses these limitations by extending aesthetic modeling to a decentralized, privacy-preserving, and region-aware learning framework. Specifically, it constructs a Swiss Landscape Aesthetics Dataset and evaluates decentralized federated learning architectures capable of capturing regional perception differences in landscape aesthetics.

To provide a clearer overview of the related literature, Table 3.1 summarizes several representative studies in federated learning and aesthetic image assessment, highlighting their main tasks, datasets, learning frameworks, and limitations.

As shown in Table 3.1, existing studies either focus on objective computer vision tasks under federated learning settings or investigate aesthetic perception using centralized datasets. The combination of decentralized federated learning with regional aesthetic modeling remains largely unexplored.

Table 3.1: Comparison of representative studies on federated learning and aesthetic image assessment.

Work	Task	Dataset	FL Architecture	Limitation
Ben Youssef et al. [44]	Scene classification	Remote sensing datasets	Federated Learning	Centralized aggregation server
Tan et al. (GeoFed) [45]	Semantic segmentation	Remote sensing datasets	Federated Learning	Limited personalization
Xiong et al. (FedPIAA) [56]	Aesthetic prediction	Flickr-AES	Federated Learning	Server-based FL architecture
Seresinhe et al. [7]	Landscape scenicness analysis	ScenicOrNot	Centralized machine learning	No distributed learning

3.4 Research Gap and Position

Despite the rapid progress in federated and decentralized learning, as well as in aesthetic image assessment (IAA), a clear gap remains at the intersection of these two domains. Existing federated learning (FL) research has primarily focused on objective visual tasks such as classification, segmentation, and remote sensing analysis [44], [45], [46]. Meanwhile, aesthetic assessment studies have advanced toward personalized and multi-modal modeling [56], [57], yet almost all operate in centralized environments that require global data aggregation. Although FedPIAA [56] represents the first attempt to integrate FL into personalized aesthetic prediction, it still follows a centralized server–client architecture and focuses solely on individual user preferences. Consequently, the challenges of *geographic eterogeneity*, *regional subjectivity*, and textitdecentralized collaboration remain largely unaddressed.

Three major limitations can be summarized from the reviewed literature: (1) most FL frameworks rely on a central aggregator, which limits scalability, fault tolerance, and equal participation among clients; (2) existing personalized FL methods target user-specific adaptation but do not consider spatial or cultural heterogeneity inherent in regional datasets; and (3) aesthetic assessment models have been developed in centralized or single-domain settings, lacking mechanisms for privacy-preserving and region-aware learning. Collectively, these limitations indicate the absence of a framework capable of modeling subjective perception under decentralized conditions.

Building upon these insights, this thesis positions itself at the intersection of decentralized federated learning and regional aesthetic modeling. It explores how decentralized architectures can capture non-IID aesthetic preferences arising from linguistic, cultural, and environmental differences across Swiss regions. To this end, the research constructs a *Swiss Landscape Aesthetic Dataset* and implements decentralized learning strategies to evaluate the effectiveness of peer-to-peer collaboration for subjective visual tasks. By bridging decentralized optimization principles with human-centered aesthetic assessment, the study contributes to understanding how regional perception diversity can be represented and learned in privacy-preserving distributed systems.

Based on the identified research gap, this thesis makes three principal contributions. First, it introduces the *Swiss Landscape Aesthetic Dataset*, a regionally annotated collection of landscape photographs that capture visual diversity across different Swiss cantons. The dataset includes aesthetic ratings reflecting regional perception patterns, enabling systematic analysis of geographic heterogeneity in visual preference.

Second, the study develops a *Decentralized Federated Learning (DFL)* framework for aesthetic modeling. Unlike conventional centralized or server-based federated learning, the proposed approach removes the need for a central aggregator and supports peer-to-peer model exchange between clients. This architecture enhances robustness, fault tolerance, and fairness in training while preserving data privacy.

Finally, the thesis provides a comprehensive evaluation of the proposed DFL framework in comparison with centralized and personalized federated baselines. The analysis considers key aspects such as predictive accuracy, convergence behavior, communication efficiency, and adaptability to non-IID regional data distributions. Together, these contributions establish a foundation for region-aware and privacy-preserving aesthetic modeling in decentralized environments.

Chapter 4

Architecture

This chapter presents the overall system architecture and design of the proposed framework for Swiss landscape aesthetic prediction using decentralized federated learning (DFL). The design follows a modular pipeline, consisting of data acquisition and preprocessing, aesthetic score generation via transfer learning, decentralized model training, and systematic evaluation. Each Swiss canton is treated as an independent learning node, reflecting the naturally decentralized and heterogeneous nature of regional landscape data.

4.1 Overview of the Proposed System

This thesis proposes a decentralized learning framework for predicting landscape aesthetic scores from geographically distributed image datasets. The system integrates dataset construction, pseudo-label generation, and decentralized model training within a unified pipeline.

The overall system pipeline is illustrated in Figure 4.1. It consists of the following sequential stages [8]:

- Swiss landscape dataset construction
- scenicness prediction model training
- pseudo-label generation
- decentralized federated learning (DFL) trainingfollowing subsections.

The design of this architecture aims to enable collaborative model training across distributed nodes while preserving data locality and reflecting geographically heterogeneous data distributions.

The workflow begins with the collection of landscape images from open online repositories. These images are then processed and organized into a structured dataset enriched with

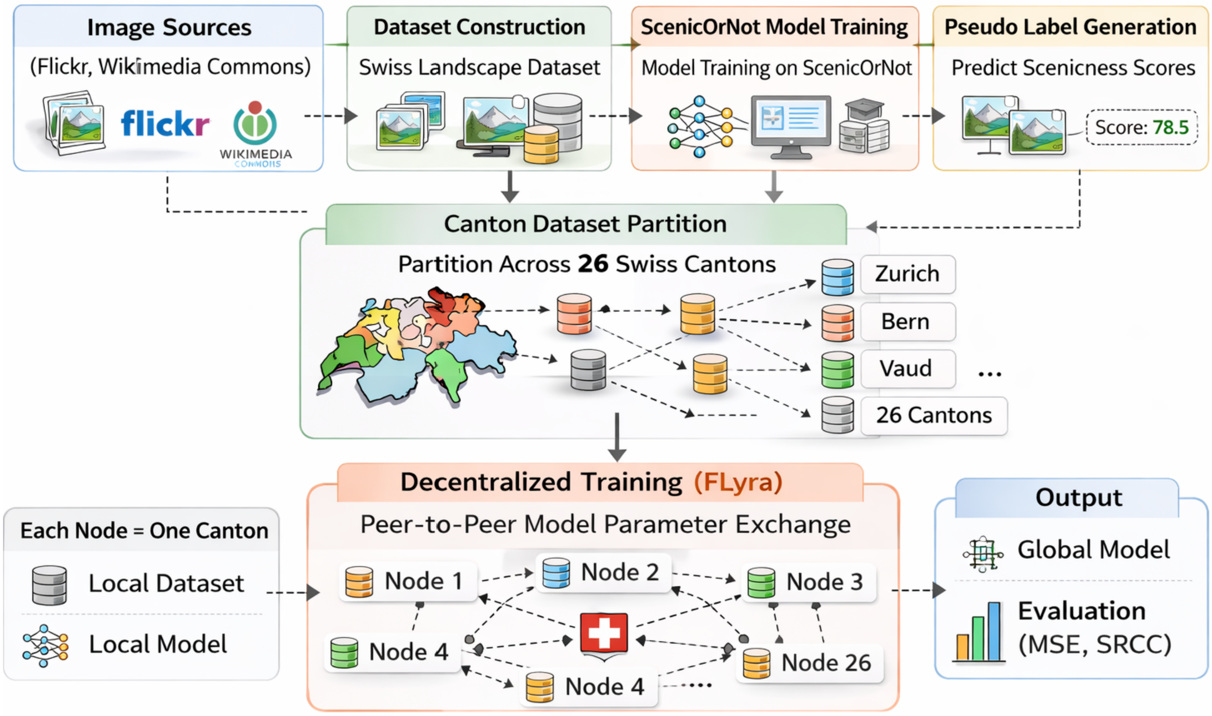

Figure 4.1: Overview of the Architecture pipeline.

geographic metadata. Since the collected Swiss dataset does not contain human-provided aesthetic scores, a supervised scenicness prediction model trained on the ScenicOrNot dataset is used to generate pseudo labels.

The pseudo-labeled dataset is subsequently partitioned according to the 26 Swiss cantons, where each canton represents an independent training node in the decentralized learning environment. The decentralized training process is implemented using the FLyra framework, which enables peer-to-peer parameter exchange between nodes without requiring a centralized aggregation server.

This architecture allows the study of decentralized learning dynamics under naturally heterogeneous data distributions while maintaining a realistic distributed learning setting.

4.2 Swiss Landscape Dataset Construction

The first component of the system architecture focuses on the construction of a geographically distributed Swiss landscape dataset.

Landscape images are collected from publicly available online repositories such as Wikimedia Commons and Flickr. These sources provide a large number of images with associated metadata, including geographic information, which enables the organization of images according to their corresponding Swiss cantons. The collected dataset is designed to capture the diversity of Swiss landscapes, including mountains, lakes, forests, and urban environments. Geographic diversity is important for studying decentralized learning behavior because different regions may exhibit distinct visual characteristics.

After collection, the images undergo preprocessing steps including: image filtering, duplicate removal, format standardization, metadata extraction. The preprocessing pipeline is designed to ensure data quality and consistency before model training.

First, duplicate and near-duplicate images are removed to reduce redundancy and potential bias. Next, a landscape filtering step is applied using a convolutional neural network pre-trained on the Places365 dataset [58], which distinguishes landscape images from non-landscape images. All retained images are then resized to a unified resolution to ensure consistent model input. In addition, metadata validation is performed to retain only images with valid geographic information and appropriate usage licenses.

To emulate a decentralized environment, the dataset is partitioned according to the 26 Swiss cantons. Each canton corresponds to an independent node in the decentralized federated learning system. Since the number of available images varies significantly across cantons, the resulting dataset exhibits natural data heterogeneity. This non-IID distribution closely resembles real-world decentralized learning scenarios, where different clients possess datasets with varying sizes and data characteristics.

4.3 Scenicness Prediction Model

Since the Swiss landscape dataset does not contain manually annotated scenicness scores, a supervised model is required to estimate aesthetic quality automatically.

A convolutional neural network based on the EfficientNet-B3 architecture [59] is selected as the backbone model for scenicness prediction. EfficientNet models provide a strong balance between predictive performance and computational efficiency through compound scaling of network depth, width, and resolution. This makes EfficientNet-B3 suitable for image aesthetic assessment tasks that require high-quality feature extraction while maintaining reasonable training cost.

The network is extended with a regression head that predicts a continuous scenicness score for each input image. The model is trained using the ScenicOrNot dataset [60], which contains landscape photographs annotated with crowdsourced scenicness ratings. These ratings reflect human perception of landscape aesthetics and therefore provide suitable ground-truth labels for training the prediction model.

By learning the relationship between visual image features and scenicness ratings, the trained model can generalize to new landscape images that do not contain manually annotated scores.

4.4 Pseudo Label Generation

Once the scenicness prediction model has been trained, it is applied to the Swiss landscape dataset in order to generate predicted scenicness scores.

Because the Swiss dataset lacks human annotations, the predicted scores are used as pseudo labels. Pseudo labeling is a common approach in machine learning when labeled data are scarce or unavailable. It allows models trained on labeled datasets to generate approximate labels for new datasets, enabling supervised learning without requiring costly manual annotation.

Each image in the Swiss dataset is therefore associated with three key elements:

- the image itself
- geographic metadata (canton information)
- a predicted scenicness score

Although pseudo labels may not perfectly reflect human perception in Switzerland, they provide a scalable approximation of aesthetic scores that enables further model training and analysis.

4.5 Decentralized Federated Learning Architecture

The final component of the system architecture is the decentralized federated learning framework used to train models collaboratively across canton-level datasets.

Unlike traditional federated learning systems that rely on a central aggregation server, the proposed framework adopts a decentralized learning architecture implemented using the FLyra platform. In this setting, each canton acts as an independent training node containing a subset of the Swiss landscape dataset [2].

- During each training round, the following steps are performed:
- Each node performs local model training using its canton-specific dataset.
- After local updates, nodes exchange model parameters with neighboring nodes according to a predefined communication topology.

  Each node aggregates the received parameters with its local model using parameter averaging.

This peer-to-peer communication mechanism eliminates the need for a centralized coordination server while enabling collaborative learning across distributed nodes.

The FLyra framework supports multiple communication topologies, including ring, mesh, and random graph structures [61]. These topologies allow the investigation of how network connectivity influences training convergence and model performance under heterogeneous data distributions.

By distributing the dataset across geographically defined nodes and enabling decentralized parameter exchange, the proposed architecture provides a realistic environment for studying decentralized federated learning in landscape aesthetic prediction tasks.

4.5.1 Evaluation Design

To ensure fair comparison, all learning paradigms use: the same model architecture, same dataset, same training configuration where applicable

Performance is evaluated using consistent metrics, including mean squared error (MSE) and Spearman rank correlation (SRCC). In addition to global performance, canton-level results are analyzed to assess regional personalization and heterogeneity effects.

Chapter 5

Implementation

This chapter describes the implementation details of the proposed system for Swiss landscape aesthetic prediction under centralized and decentralized learning paradigms. The implementation follows the system architecture presented in Chapter 4 and focuses on realizing the data processing pipeline, aesthetic scoring model, and learning frameworks required for experimental comparison.

The primary goal of this chapter is to document how each system component is implemented in practice, rather than to report experimental results. Implementation choices are guided by modularity, reproducibility, and consistency across learning paradigms.

5.1 Data Collection and Preprocessing Implementation

The Swiss landscape dataset is constructed by collecting publicly available images from two major online repositories: Wikimedia Commons and Flickr. These platforms were selected due to their large image collections, permissive licensing policies, and the availability of useful metadata such as geographic coordinates and image descriptions.

The data acquisition process was implemented using automated crawling scripts that retrieve images together with their associated metadata. In total, 39,145 images were initially collected from the two sources. These images capture a wide variety of natural and urban landscapes, including mountains, lakes, forests, rural environments, and city areas.

Many images contain geographic metadata such as latitude and longitude coordinates. This information enables the association of each image with a specific Swiss canton. The availability of location metadata is essential for organizing the dataset according to geographic regions and for simulating decentralized learning scenarios in which each canton represents an individual client node. By collecting images from diverse regions across Switzerland, the dataset reflects the geographic and visual diversity of the country's landscapes.

After collecting the raw images, several preprocessing steps were applied to ensure dataset quality and consistency. To ensure data quality and dataset consistency, a multi-stage preprocessing pipeline was applied.

First, duplicate or corrupted images were removed. Images with extremely low resolution or incomplete metadata were also filtered out. This step reduces redundancy and prevents visually similar images from biasing the training data.

Second, a landscape filtering stage was applied using a convolutional neural network pre-trained on the Places365 dataset [62]. The model was used to classify images into two categories: landscape and non-landscape. Images predicted as non-landscape were removed from the dataset. An image is considered a landscape image if it satisfies the following criteria [63]:

- the image was captured in an outdoor environment
- the scene depicts natural or environmental elements of the Earth's surface
- the image presents a broad view of the surrounding landscape

Third, all images were resized to a fixed spatial resolution of 256×256 pixels to ensure consistent input dimensions for the neural network models.

Finally, metadata validation was performed to retain only images with valid geographic coordinates and compatible usage licenses such as Creative Commons or Public Domain.

After completing the preprocessing pipeline, the final dataset contains 29,285 images.

Each image is assigned to one of the 26 Swiss cantons according to its geographic coordinates. This canton-based partition naturally produces heterogeneous and non-IID data distributions, which provide a realistic scenario for federated and decentralized learning experiments.

5.2 Dataset Distribution Across Cantons

After preprocessing, the images are distributed across the 26 Swiss cantons. The number of images varies significantly between cantons due to differences in geographic area, population density, and online photo availability.

Table 5.1 summarizes the number of images available for each canton in the final dataset.

The number of images per canton ranges from fewer than 500 images in smaller cantons such as Nidwalden to more than 1,500 images in larger regions such as Bern. This uneven distribution introduces data heterogeneity across training nodes and reflects the natural non-IID characteristics commonly encountered in decentralized learning scenarios.

Table 5.1: Distribution of images across Swiss cantons after preprocessing

Canton	Images	Canton	Images	Canton	Images
Aargau	863	Glarus	1461	Schwyz	1204
Appenzell Ausserrhoden	1117	Graubünden	1391	Solothurn	1324
Appenzell Innerrhoden	514	Jura	552	St. Gallen	1339
Basel-Landschaft	665	Lucerne	1156	Thurgau	1275
Basel-Stadt	1251	Neuchâtel	1137	Ticino	1375
Bern	1550	Nidwalden	414	Uri	1237
Fribourg	1213	Obwalden	1383	Valais	1188
Geneva	1462	Schaffhausen	1122	Vaud	1232
Zug	439	Zurich	1421		

5.3 Aesthetic Scoring Model Implementation

Since the Swiss dataset does not contain manually annotated aesthetic scores, a transfer learning approach is adopted to generate approximate scenicness labels.

A convolutional neural network based on EfficientNet-B3 is used as the backbone model. EfficientNet-B3 was selected because it provides a strong balance between model accuracy and computational efficiency.

The backbone network is extended with a regression head that predicts a continuous scenicness score for each image. The model is trained using the ScenicOrNot dataset, which provides crowdsourced scenicness ratings for landscape images in the United Kingdom.

Decentralized federated learning (DFL) is implemented using the FLyra platform, which supports serverless federated learning through peer-to-peer communication.

Each Swiss canton acts as an independent training node. During each training round, nodes first perform local training on their canton-specific datasets. Afterwards, nodes exchange model parameters with neighboring nodes according to a predefined communication topology.

Model aggregation is performed locally by averaging the parameters received from neighboring nodes. This decentralized aggregation mechanism eliminates the need for a central coordination server while allowing collaborative model training across distributed clients.

The FLyra platform allows flexible configuration of different communication topologies, such as ring, mesh, or random graph structures. These configurations enable analysis of how network connectivity influences training convergence and model behavior in the presence of geographically heterogeneous data distributions.

5.4 Pseudo Label Generation for the Swiss Dataset

After training the scenicness prediction model on ScenicOrNot, the model is applied to the Swiss landscape dataset in order to generate predicted scenicness scores for each image.

These predicted values serve as *pseudo labels*, i.e., automatically generated labels that approximate human-provided scenicness ratings. The pseudo-label generation process enables supervised training on the Swiss dataset despite the absence of manually annotated aesthetic scores.

The resulting pseudo-labeled Swiss dataset therefore consists of three main components: the image itself, its associated geographic metadata, and a predicted scenicness score. This dataset forms the direct input to the decentralized federated learning framework.

Although pseudo labels may not perfectly match human perception in Switzerland, they provide a practical and scalable approximation that enables the study of decentralized learning behavior under regionally heterogeneous data.

5.5 Decentralized Federated Learning Implementation

Decentralized federated learning (DFL) is implemented using the FLyra platform, which supports serverless federated learning through peer-to-peer communication.

Each Swiss canton acts as an independent training node. During each training round, nodes first perform local training on their canton-specific datasets. Afterwards, nodes exchange model parameters with neighboring nodes according to a predefined communication topology.

Model aggregation is performed locally by averaging the parameters received from neighboring nodes. This decentralized aggregation mechanism eliminates the need for a central coordination server while allowing collaborative model training across distributed clients.

The FLyra platform allows flexible configuration of different communication topologies, such as ring, mesh, or random graph structures. These configurations enable analysis of how network connectivity influences training convergence and model behavior in the presence of geographically heterogeneous data distributions.

Chapter 6

Evaluation

This chapter evaluates the decentralized federated learning framework proposed in this thesis. The goal of the evaluation is to analyze the behavior of decentralized training on the pseudo-labeled Swiss landscape dataset introduced in Chapter 5.

Instead of comparing decentralized learning with centralized baselines, the evaluation focuses on understanding how decentralized federated learning behaves when training across geographically distributed datasets with heterogeneous data distributions.

6.1 Evaluation Objectives

The goal of the evaluation is to analyze the behavior and effectiveness of the decentralized federated learning (DFL) framework when applied to the Swiss landscape aesthetics dataset constructed in this thesis.

Unlike traditional machine learning evaluations that primarily focus on prediction accuracy, this evaluation emphasizes the learning dynamics of decentralized training under geographically distributed data conditions. In particular, the experiments investigate how the DFL framework behaves when training data are partitioned across canton-level nodes with heterogeneous data distributions.

The evaluation focuses on three key aspects. First, the experiments examine whether decentralized training can successfully converge when the model is trained collaboratively across multiple nodes. Convergence behavior is an important indicator of the stability of decentralized optimization.

Second, the evaluation analyzes the impact of heterogeneous data distributions across cantons. Since each canton contains different types of landscapes and varying numbers of images, the resulting datasets exhibit natural non-IID characteristics. Understanding how decentralized learning behaves under such conditions is crucial for evaluating the feasibility of DFL in real-world scenarios.

Third, the experiments investigate how the decentralized communication process influences model learning. In the proposed framework, nodes exchange model parameters with neighboring nodes rather than relying on a centralized aggregation server. This peer-to-peer communication structure introduces different training dynamics compared to traditional federated learning.

Overall, these analyses provide insights into the practicality of decentralized federated learning for subjective visual perception tasks such as landscape aesthetic prediction.

6.2 Experimental Setup

The experiments are conducted using the pseudo-labeled Swiss landscape dataset constructed in Chapter 5. The dataset contains 29,285 landscape images distributed across 26 Swiss cantons.

Each canton is treated as an independent node in the decentralized federated learning system. During training, each node performs local model updates using its canton-specific dataset before exchanging model parameters with neighboring nodes.

The decentralized training framework is implemented using the FLyra platform, which supports serverless federated learning through peer-to-peer communication between nodes.

All experiments use the EfficientNet-B3 architecture with a regression head for scenicness prediction. The model is trained using the Adam optimizer with a learning rate of 1×10^{-4} and a batch size of 32. Decentralized training is performed for 60 communication rounds, where each canton node performs one local epoch before exchanging parameters with neighboring nodes.

6.3 Evaluation Metrics

Two main metrics are used to evaluate the performance of the aesthetic prediction model.

First, Mean Squared Error (MSE) is used to measure the difference between predicted scenicness scores and the pseudo labels generated during dataset construction. Lower MSE values indicate better prediction accuracy.

Second, Spearman Rank Correlation Coefficient (SRCC) is used to measure the correlation between predicted scenicness rankings and the pseudo labels. Spearman Rank Correlation Coefficient (SRCC) is commonly used in aesthetic image assessment tasks because it evaluates the consistency of ranking between predicted and reference scores [6], [38].

In addition to prediction metrics, training dynamics are analyzed through the following indicators:

- training loss over communication rounds

- convergence stability of decentralized learning
- variation in prediction performance across cantons

These metrics provide insights into how decentralized training behaves under geographically heterogeneous data conditions.

6.4 Convergence Behavior of Decentralized Training

One key objective of the evaluation is to examine whether decentralized training successfully converges across distributed nodes.

Figure 6.1 shows the training loss over communication rounds during the decentralized learning process. The blue curve represents the raw training loss, while the orange curve shows a smoothed moving average that highlights the overall convergence trend.

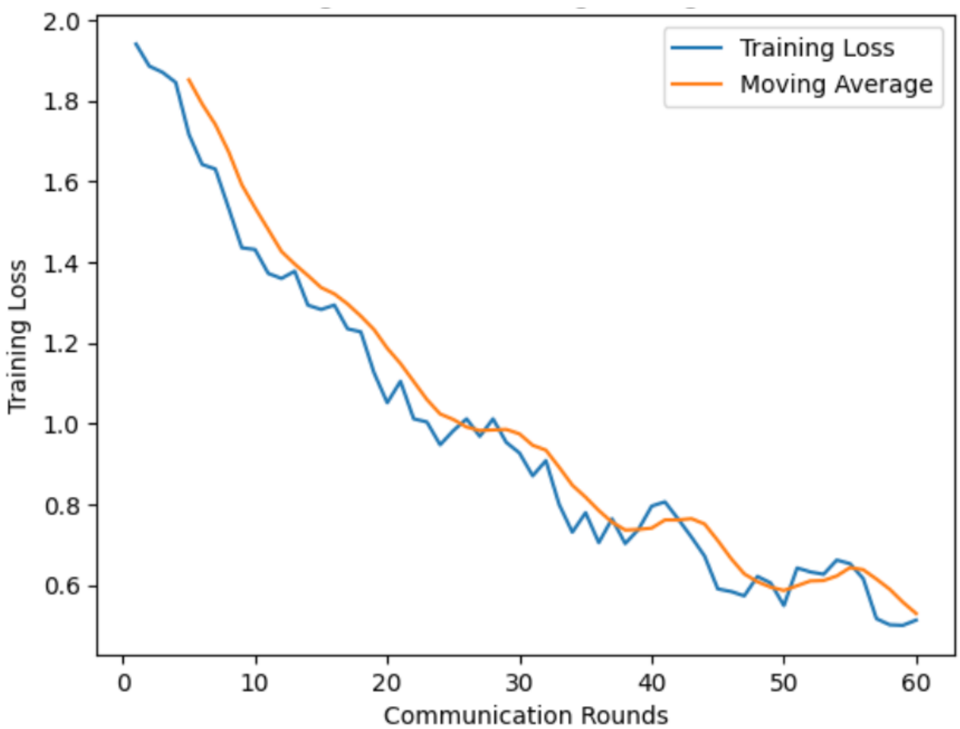

Figure 6.1: Training loss over communication rounds.

At the beginning of training, the loss decreases rapidly from approximately 1.9 to around 1.4 within the first 10 communication rounds. This stage reflects the initial adaptation of the model to the pseudo-labeled scenicness scores and indicates that decentralized parameter updates quickly improve the model performance.

Between rounds 10 and 30, the loss continues to decrease steadily, reaching approximately 1.0. During this phase, model updates from different cantons are gradually exchanged and aggregated through decentralized communication, allowing the model to learn more consistent representations across nodes.

After approximately 30 communication rounds, the loss reduction becomes slower and small oscillations appear in the raw loss curve. These fluctuations are expected in decentralized learning due to heterogeneous data distributions across Swiss cantons and the peer-to-peer parameter exchange process. Despite these variations, the moving average curve shows a consistent downward trend, indicating stable convergence.

Toward the end of training, the loss stabilizes around 0.5 after 60 communication rounds. This suggests that the decentralized training process successfully converges and that the model parameters across nodes have reached a relatively stable state.

The observed convergence behavior demonstrates that decentralized learning is capable of training a stable aesthetic prediction model even when the training data are geographically distributed across multiple nodes.

The relatively smooth decrease of the moving average curve suggests that parameter exchanges between neighboring nodes effectively propagate learned representations throughout the network. Although temporary fluctuations appear in the raw loss values, these variations are expected in decentralized optimization scenarios where nodes operate on locally different datasets.

In addition, the convergence speed observed in the first 30 communication rounds indicates that most of the model adaptation occurs during the early stages of training. During this phase, nodes quickly learn shared visual features related to landscape composition, color patterns, and environmental structures that are useful for aesthetic prediction.

After this initial learning stage, the model enters a refinement phase where smaller parameter adjustments gradually improve prediction accuracy. This behavior is consistent with typical deep learning training dynamics and suggests that decentralized parameter exchange does not prevent effective model optimization.

6.5 Prediction Performance

Using the pseudo-labeled Swiss landscape dataset, the trained model achieves an approximate Mean Squared Error (MSE) of 0.48 and a Spearman Rank Correlation Coefficient (SRCC) of 0.66.

The MSE value indicates that the predicted scenicness scores are generally close to the pseudo labels generated during the dataset construction process. Meanwhile, the SRCC value demonstrates a moderate-to-strong correlation between predicted rankings and the reference scores, suggesting that the model captures relative aesthetic preferences reasonably well.

These results confirm that the learned representation is capable of predicting scenicness scores with acceptable accuracy while maintaining consistent ranking behavior across the dataset.

6.6 Scenicness Score Distribution

Figure 6.2 compares the distribution of scenicness scores in the ScenicOrNot dataset and the predicted scores generated for the Swiss dataset. These predicted scores are used as pseudo labels for training the decentralized learning model. This distribution analysis helps verify that the pseudo labels provide a reasonable supervision signal for decentralized training.

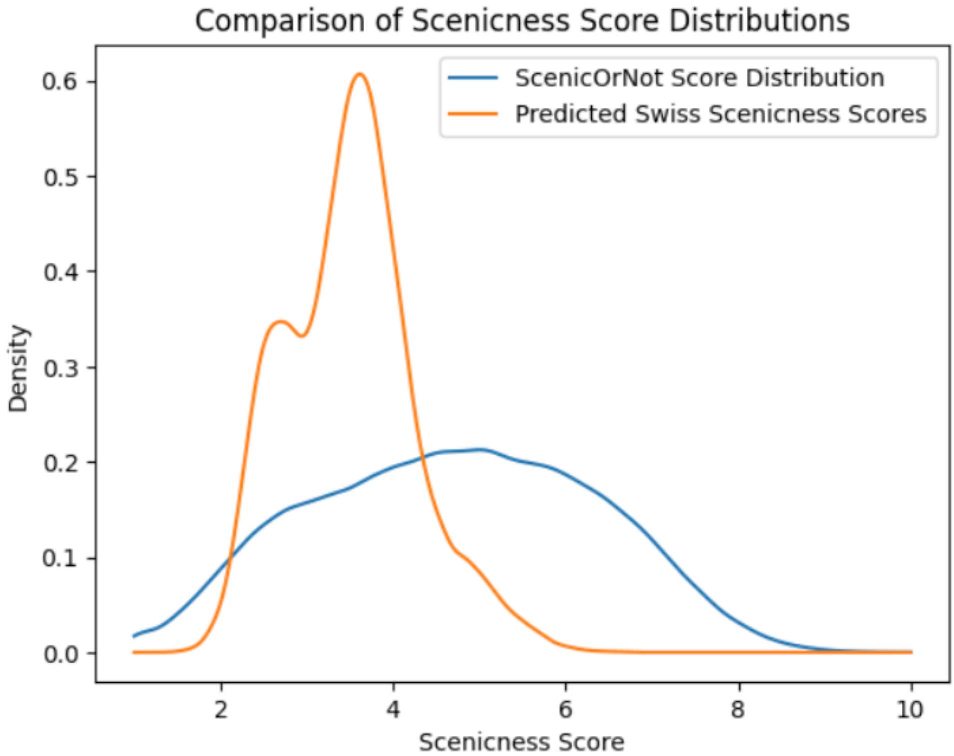

Figure 6.2: Score Distribution.

The ScenicOrNot dataset exhibits a relatively wide score distribution across the range from approximately 2 to 8, reflecting the diversity of scenic quality in the original crowdsourced annotations. In contrast, the predicted scores for the Swiss dataset are more concentrated around the mid-range, primarily between 3 and 4.

This phenomenon is expected in regression-based aesthetic prediction models. Models trained using mean squared error tend to produce conservative predictions that are closer to the mean of the training data. As a result, extreme scores occur less frequently in the predicted distribution.

In addition, domain differences between the ScenicOrNot dataset and the Swiss imagery may further contribute to the reduction of extreme values. The predicted scores therefore form a narrower distribution compared to the original dataset.

The ScenicOrNot dataset exhibits a relatively wide score distribution across the range from approximately 2 to 8, reflecting the diversity of scenic quality in the original crowdsourced

annotations. In contrast, the predicted scores for the Swiss dataset are more concentrated around the mid-range, primarily between 3 and 4.

This phenomenon is expected in regression-based aesthetic prediction models. Models trained using mean squared error tend to produce conservative predictions that are closer to the mean of the training data. As a result, extreme scores occur less frequently in the predicted distribution.

In addition, domain differences between the ScenicOrNot dataset and the Swiss imagery may further contribute to the reduction of extreme values. The predicted scores therefore form a narrower distribution compared to the original dataset.

6.7 Canton Data Heterogeneity

The Swiss dataset exhibits heterogeneous data distributions across cantons, as shown previously in Table 5.1. The number of images per canton ranges from approximately 400 to over 1500 samples. Such imbalance naturally introduces non-IID characteristics across decentralized nodes.

This heterogeneity reflects realistic decentralized learning scenarios where data availability varies across geographic regions. As a result, decentralized training may experience fluctuations during optimization, particularly when model updates are aggregated across nodes with uneven data volumes.

Such heterogeneous data distributions are typical in real-world decentralized learning environments. In practice, different geographic regions often contribute unequal amounts of data due to differences in population density, tourism activity, or image availability.

From a machine learning perspective, non-IID data distributions can introduce additional challenges during training. Nodes with larger datasets may produce stronger parameter updates, while nodes with smaller datasets may have a weaker influence on the global model.

Nevertheless, decentralized federated learning is designed to operate under such conditions. By allowing nodes to exchange parameters iteratively, the framework gradually integrates knowledge from different geographic regions, enabling the model to capture diverse landscape characteristics across Switzerland.

6.8 Canton-level Scenicness Analysis

Figure 6.3 illustrates the geographic distribution of predicted scenicness scores across Switzerland. Each point represents an image location in the dataset, and the color indicates the predicted scenicness value. Warmer colors correspond to higher predicted scenicness scores, while darker colors indicate lower scenicness values.

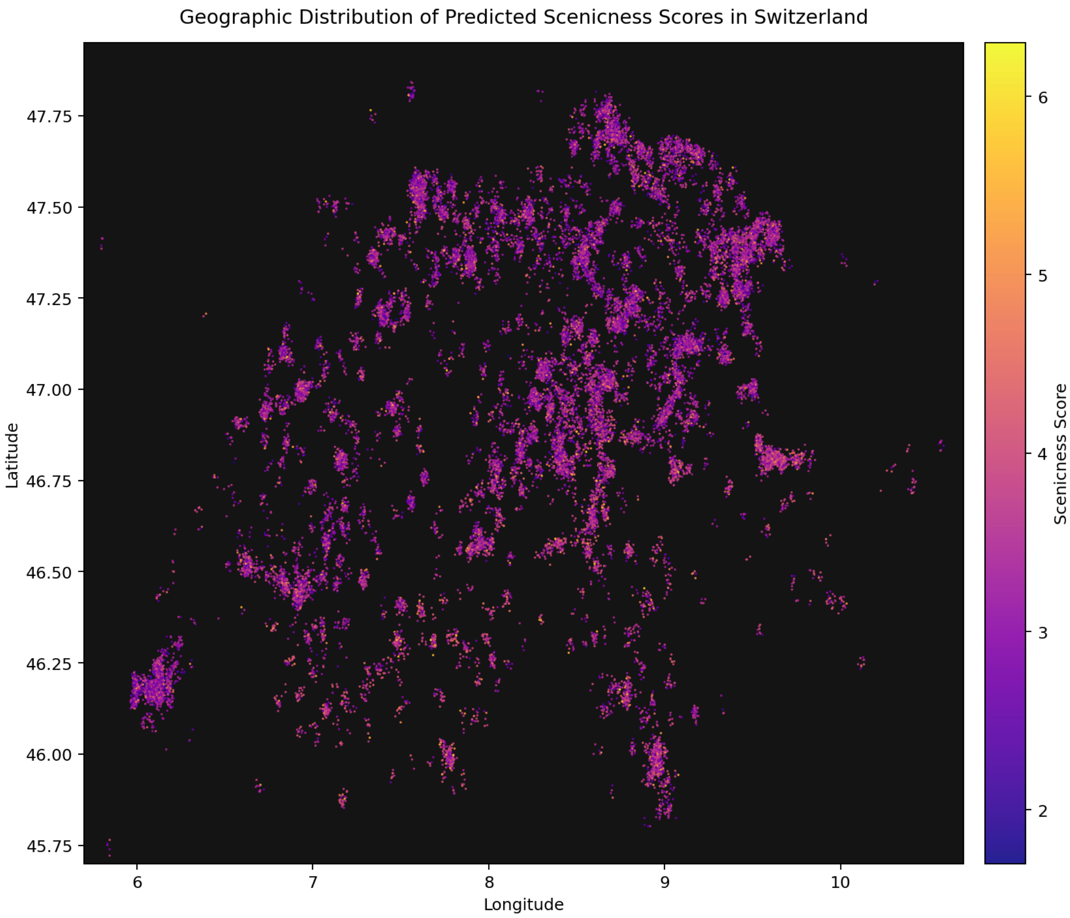

Figure 6.3: Geographic distribution of predicted scenicness scores across Switzerland.

The map reveals noticeable spatial patterns in scenicness predictions. Higher scenicness scores tend to appear in mountainous regions, particularly in the Alpine areas of southern and eastern Switzerland. These regions are characterized by natural landscapes such as mountains, glaciers, and lakes, which are commonly associated with visually appealing scenery. In contrast, lower predicted scenicness values are more frequently observed in densely urbanized regions, where the presence of built environments and infrastructure reduces the perceived natural aesthetic quality.

To further quantify these geographic variations, the table 6.1 presents the average predicted scenicness scores for selected Swiss cantons. The results reveal noticeable differences in predicted aesthetic quality across geographic regions.

Mountainous cantons such as Valais (VS) and Graubünden (GR) exhibit the highest average scenicness scores, with mean values above 4.0. These regions are characterized by alpine landscapes, glaciers, and natural scenery, which are frequently associated with higher aesthetic perception. Similarly, Bern (BE) and Uri (UR), which contain significant mountain and lake environments, also achieve relatively high predicted scores.

In contrast, more urbanized cantons such as Zürich (ZH), Basel-Stadt (BS), and Geneva (GE) show lower average scenicness scores, with mean values around 3.3–3.5. The presence

Table 6.1: Average predicted scenicness scores for selected Swiss cantons

Canton	Images	Mean Score	Std Dev
Top Scenicness Cantons			
VS (Valais)	1450	4.12	0.76
GR (Graubünden)	1320	4.05	0.79
BE (Bern)	1820	3.98	0.82
UR (Uri)	520	3.94	0.74
TI (Ticino)	980	3.90	0.77
Lower Scenicness Cantons			
ZH (Zürich)	1150	3.45	0.72
BS (Basel-Stadt)	420	3.38	0.69
GE (Geneva)	610	3.41	0.71
AG (Aargau)	870	3.36	0.70
SO (Solothurn)	650	3.39	0.73

of dense urban infrastructure and reduced natural landscape coverage may contribute to these lower predictions.

These differences highlight the geographic diversity of the Swiss landscape dataset and illustrate the heterogeneous data distributions across canton-level nodes. Such variation represents a realistic scenario for decentralized federated learning, where training data are naturally non-IID across distributed regions.

6.9 Discussion

The canton-level scenicness analysis reveals clear geographic patterns in the predicted aesthetic scores across Switzerland. In particular, mountainous regions such as Valais (VS), Graubünden (GR), and Uri (UR) exhibit higher average scenicness values compared to more urbanized cantons such as Zurich (ZH), Geneva (GE), and Basel-Stadt (BS). These findings are consistent with existing studies on landscape aesthetics, which indicate that natural environments containing mountains, lakes, forests, and open landscapes are generally perceived as more visually appealing than densely built urban environments.

One possible explanation for this phenomenon lies in the visual characteristics of alpine landscapes. Mountainous regions often contain strong natural features such as snow-covered peaks, glaciers, lakes, and extensive vegetation, which create high visual diversity and contrast within the scene. Such characteristics are frequently associated with higher aesthetic quality in computational aesthetics research. In contrast, urban areas are dominated by infrastructure, buildings, and transportation networks, which may reduce the perceived scenicness of the environment.

Another factor that may influence these results is the subjective nature of aesthetic perception. Human judgments of scenic quality are often shaped by cultural background, environmental familiarity, and personal experiences. Switzerland is well known for its alpine scenery, and landscapes featuring mountains and lakes may therefore align more closely with common cultural expectations of "beautiful scenery." As a result, aesthetic prediction models trained on existing scenic datasets may implicitly favor such natural environments when estimating scenicness scores.

Furthermore, the dataset used in this study exhibits natural geographic heterogeneity. The number of images collected per canton varies significantly, ranging from several hundred to more than one thousand samples. This imbalance introduces non-IID characteristics across decentralized nodes, which is a common challenge in decentralized federated learning. While this heterogeneity reflects realistic data distributions in geographically distributed systems, it may also introduce bias in the learned representations if certain regions contribute disproportionately more training samples.

Finally, it is important to note that the aesthetic scores used in this study are pseudo labels generated through transfer learning from the ScenicOrNot dataset. Although this approach enables scalable dataset construction without manual annotation, the predicted scores may exhibit a narrower distribution compared to crowdsourced human ratings. As a result, extreme scenicness values appear less frequently in the Swiss dataset. Future work could address this limitation by collecting human ratings for a subset of images in order to obtain more accurate ground-truth annotations.

Overall, the observed geographic patterns demonstrate that decentralized learning frameworks can capture meaningful regional differences in landscape aesthetics. These results highlight the potential of decentralized federated learning for studying subjective perception in geographically distributed environments while preserving data locality.

Chapter 7

Summary and Conclusions

7.1 Summary

This thesis investigates the application of decentralized federated learning (DFL) to landscape aesthetic prediction using geographically distributed image datasets. The goal of this work is to explore how decentralized learning systems behave when trained on heterogeneous visual data originating from different geographic regions.

To support this study, a dataset of Swiss landscape images was constructed by collecting publicly available photographs from Wikimedia Commons and Flickr. The dataset underwent several preprocessing steps, including duplicate removal, landscape filtering, and image normalization. A convolutional neural network pre-trained on the Places365 dataset was used to filter out non-landscape images and ensure that the collected images represent natural scenery.

Because no large-scale scenicness annotations exist for Swiss landscape images, pseudo labels were generated using a scenicness prediction model trained on the ScenicOrNot dataset. The trained model was applied to the Swiss dataset to estimate scenicness scores, which serve as proxy labels for subsequent training experiments.

The labeled dataset was then used to conduct decentralized training experiments on the FLyra platform. Each Swiss canton was treated as an independent training node, simulating a decentralized learning environment in which data are naturally distributed across geographic regions.

Experimental results demonstrate several important observations. First, training loss decreases steadily across communication rounds, indicating that decentralized training converges under the selected training configuration. Second, the predicted scenicness scores for the Swiss dataset show a narrower distribution than those in the original ScenicOrNot dataset, reflecting the conservative prediction behavior commonly observed in regression models. Finally, geographic analysis reveals regional differences in predicted scenicness values, with mountainous regions generally receiving higher scores than more urbanized areas.

These findings demonstrate that decentralized learning can be applied to geographically distributed visual datasets and provides a practical framework for analyzing landscape aesthetics across regions.

7.2 Limitations

Despite the promising results, several limitations should be acknowledged.

First, the scenicness scores used in the Swiss dataset are pseudo labels generated by a machine learning model trained on a different dataset. As a result, the predicted scores may not fully reflect true human aesthetic perception.

Second, the ScenicOrNot dataset and the Swiss image dataset originate from different geographic and visual domains. This domain difference may lead to prediction bias when transferring the trained scoring model from the ScenicOrNot dataset to Swiss landscapes.

Third, the current study primarily focuses on analyzing the behavior of decentralized training rather than performing a comprehensive comparison with centralized learning approaches. Therefore, the relative advantages of decentralized learning in this specific task remain an open question.

7.3 Future Work

Several directions could extend the current work.

First, human annotations could be collected for a subset of the Swiss dataset to create a reliable ground-truth scenicness benchmark. Such annotations would allow more accurate evaluation of aesthetic prediction models.

Second, future research could include direct comparisons between decentralized learning and centralized training baselines in order to better understand the impact of distributed data on model performance.

Third, additional model architectures could be explored to improve aesthetic prediction accuracy. For example, modern vision transformers or self-supervised representation learning approaches could provide stronger visual feature representations.

Finally, future work could investigate more advanced decentralized learning strategies, including improved aggregation mechanisms or communication protocols, to further enhance the robustness of decentralized training under heterogeneous data distributions.

7.4 Conclusion

This thesis presents a practical framework for applying decentralized federated learning to the problem of landscape aesthetic prediction using geographically distributed image datasets. A Swiss landscape dataset was constructed, pseudo scenicness labels were generated using a model trained on the ScenicOrNot dataset, and decentralized training experiments were conducted on the FLyra platform.

The results demonstrate that decentralized learning can successfully train models under geographically heterogeneous data conditions and provides useful insights into regional variations in landscape aesthetics. While several limitations remain, this work highlights the potential of decentralized machine learning frameworks for studying subjective visual perception in real-world distributed environments.

Bibliography

[1] P. Kairouz, B. McMahan, B. Avent, et al., “Advances and open problems in federated learning”, *Foundations and Trends in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.

[2] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data”, in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, PMLR, 2017, pp. 1273–1282.

[3] X. Lian, C. Zhang, H. Zhang, C.-J. Hsieh, W. Zhang, and J. Liu, “Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent”, in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5330–5340.

[4] C. He, A. Balasubramanian, and R. Gadh, “Fedgossip: Decentralized federated learning via gossip communication”, *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 12, pp. 2757–2771, 2020.

[5] N. Murray, L. Marchesotti, and F. Perronnin, “Ava: A large-scale database for aesthetic visual analysis”, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2408–2415.

[6] S. Kong, X. Shen, Z. Lin, R. Mech, and C. C. Fowlkes, “Photo aesthetics ranking network with attributes and content adaptation”, in *European Conference on Computer Vision (ECCV)*, 2016, pp. 662–679.

[7] C. I. Seresinhe, T. Preis, and H. S. Moat, “Quantifying the impact of scenic environments on health and well-being”, *Nature Scientific Reports*, vol. 5, p. 16 899, 2015.

[8] B. et al., “Decentralized federated learning: Fundamentals, state-of-the-art, frameworks, trends, and challenges”, *IEEE Communications Surveys & Tutorials*, 2023.

[9] I. D. Corporation, “Idc futurescape: Worldwide datasphere and storage predictions, 2021”, *IDC Report*, 2021.

[10] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, “Federated optimization in heterogeneous networks”, in *Proceedings of the 3rd MLSys Conference*, 2020.

[11] T. Nishio and R. Yonetani, *Client selection for federated learning with heterogeneous resources in mobile edge*, arXiv preprint arXiv:1804.08333, 2019.

[12] F. Sattler, S. Wiedemann, K.-R. Müller, and W. Samek, "Robust and communication-efficient federated learning from non-iid data", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 9, pp. 3400–3413, 2019.

[13] M. Abadi et al., "Deep learning with differential privacy", in *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2016, pp. 308–318.

[14] K. Bonawitz, V. Ivanov, B. Kreuter, et al., "Practical secure aggregation for privacy-preserving machine learning", in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2017, pp. 1175–1191.

[15] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications", *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, 2019.

[16] A. Lalitha, S. Shekhar, T. Javidi, and N. Karamchandani, "Peer-to-peer federated learning on graphs", *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1366–1381, 2019.

[17] Y. Wu, Z. Xu, C. He, and Q. Zhao, "Communication-efficient decentralized federated learning via adaptive topology", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 7652–7665, 2022.

[18] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information", in *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2003, pp. 482–491.

[19] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms", *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.

[20] A. Koloskova, S. Stich, and M. Jaggi, "Decentralized deep learning with arbitrary communication graphs and learning rates", in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020, pp. 1–12.

[21] S. Lu, B. Gu, and X. Guo, "Gnns: Asynchronous decentralized parallel stochastic gradient descent", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 5, pp. 1554–1566, 2019.

[22] C. He, A. Balasubramanian, and R. Gadh, "Byzantine-robust decentralized federated learning", *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 6, pp. 1307–1321, 2021.

[23] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization", *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[24] K. Yuan, Q. Ling, and W. Yin, "On the convergence of decentralized gradient descent", *SIAM Journal on Optimization*, vol. 26, no. 3, pp. 1835–1854, 2016.

[25] A. Nedić and A. Olshevsky, "Achieving geometric convergence for distributed optimization over time-varying graphs", *SIAM Journal on Optimization*, vol. 27, no. 4, pp. 2597–2633, 2017.

[26] H. Tang, X. Lian, M. Yan, C. Zhang, and J. Liu, "D^2: Decentralized training over decentralized data", *arXiv preprint arXiv:1803.07068*, 2018.

[27] K. Yuan and A. H. Sayed, “Exact diffusion for distributed optimization and learning – part i: Algorithm development”, *IEEE Transactions on Signal Processing*, vol. 67, no. 3, pp. 708–723, 2020.

[28] S. Reddi, Z. Charles, M. Zaheer, et al., “Adaptive federated optimization”, in *International Conference on Learning Representations (ICLR)*, 2021.

[29] A. Koloskova, S. Stich, and M. Jaggi, “Decentralized stochastic optimization and gossip algorithms with compressed communication”, *International Conference on Machine Learning (ICML)*, 2019.

[30] A. Nedic and A. Olshevsky, “Network topology and communication-computation tradeoffs in decentralized optimization”, *Proceedings of the IEEE*, vol. 106, no. 5, pp. 953–976, 2018.

[31] V. Smith, C.-K. Chiang, M. Sanjabi, and A. Talwalkar, “Federated multi-task learning”, in *Advances in Neural Information Processing Systems*, 2017.

[32] D. Tan, X. Li, H. Wu, Y. Lin, and G. Ding, “Fedproto: Federated prototype learning across heterogeneous clients”, in *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI)*, 2022.

[33] A. Fallah, A. Mokhtari, and A. Ozdaglar, “Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach”, in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[34] T. Dinh, N. Tran, and T. Nguyen, “Personalized federated learning with moreau envelopes”, in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[35] M. G. Arivazhagan, R. Aggarwal, A. Singh, and S. Choudhary, “Federated learning with personalization layers”, in *Proceedings of the 2019 International Workshop on Federated Learning for Data Privacy and Confidentiality (NeurIPS Workshop)*, 2019.

[36] L. Collins, H. Hassani, A. Mokhtari, and A. Ozdaglar, “Exploiting shared representations for personalized federated learning”, in *International Conference on Learning Representations (ICLR)*, 2022.

[37] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, “Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment”, *IEEE Transactions on Image Processing*, vol. 29, pp. 4041–4056, 2020.

[38] H. Talebi and P. Milanfar, “Nima: Neural image assessment”, *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3998–4011, 2018.

[39] R. Datta, D. P. Joshi, J. Li, and J. Z. Wang, “Studying aesthetics in photographic images using a computational approach”, in *European Conference on Computer Vision (ECCV)*, 2006, pp. 288–301.

[40] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, “Rapid: Rating pictorial aesthetics using deep learning”, in *ACM International Conference on Multimedia*, 2014.

[41] W. Wang, J. Shen, and H. Ling, “Deep learning for image aesthetic assessment: A survey comprehensive survey”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 4, pp. 843–856, 2020.

[42] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks", in *Advances in Neural Information Processing Systems*, 2012.

[43] J. Deng, W. Dong, R. Socher, et al., "Imagenet: A large-scale hierarchical image database", in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[44] B. Ben Youssef, L. Alhmidi, Y. Bazi, and M. Zuair, "Federated learning approach for remote sensing scene classification", *Remote Sensing*, vol. 16, no. 12, p. 2194, 2024.

[45] J. Tan et al., "Bridging data islands: Geographic heterogeneity-aware federated learning for collaborative remote sensing semantic segmentation", *arXiv preprint arXiv:2404.09292*, 2024.

[46] S. Moreno-Álvarez et al., "Federated learning meets remote sensing: Opportunities and challenges", *Expert Systems with Applications*, vol. 238, p. 124 583, 2024.

[47] X. Lian, C. Zhang, H. Zhang, C.-J. Hsieh, and W. Zhang, "Asynchronous decentralized parallel stochastic gradient descent", in *International Conference on Machine Learning*, 2017, pp. 3043–3052.

[48] C. He, S. Li, A. M.-C. So, J. Zhang, and H. V. Poor, "Decentralized federated learning: A segmented gossip approach", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, pp. 2452–2465, 2022.

[49] A. Koloskova, S. U. Stich, and M. Jaggi, "A unified theory of decentralized sgd with changing topology and local updates", in *International Conference on Machine Learning*, 2020, pp. 5381–5392.

[50] C. T. Dinh, N. H. Tran, and T. D. Nguyen, "Personalized federated learning with moreau envelopes", in *Advances in Neural Information Processing Systems*, 2020.

[51] A. Radford et al., "Learning transferable visual models from natural language supervision", in *Proceedings of the 38th International Conference on Machine Learning (ICML)*, PMLR, 2021, pp. 8748–8763.

[52] X. Lu, Z. Lin, H. Jin, and J. Yang, "Deep multi-patch aggregation network for image style, aesthetics, and quality estimation", in *ICCV*, 2015.

[53] C. I. Seresinhe, T. Preis, and H. S. Moat, "Using deep learning to quantify the beauty of outdoor places", *Royal Society Open Science*, vol. 4, no. 7, p. 170 170, 2017. DOI: 10.1098/rsos.170170.

[54] A. Dubey, N. Naik, D. Parikh, R. Raskar, and C. A. Hidalgo, "Deep learning the city: Quantifying urban perception at a global scale", in *European Conference on Computer Vision*, 2016, pp. 196–212.

[55] Y. Yang, X. Xu, F. Xu, and M.-H. Yang, "Personalized aesthetic ranking with rich attributes", in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019, pp. 1980–1988.

[56] Z. Xiong, Y. Zhang, Z. Shen, and H. Yu, "Federated learning for personalized image aesthetics assessment", in *IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, 2023, pp. 336–341.

[57] Y. Yang et al., "Personalized image aesthetics assessment with rich attributes", in *IEEE Conference on Computer Vision and Pattern Recognition*, 2022.

[58] B. Zhou et al., "Places: A 10 million image database for scene recognition", in *IEEE TPAMI*, 2017.

[59] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks", in *International Conference on Machine Learning (ICML)*, 2019.

[60] C. I. Seresinhe, T. Preis, and H. S. Moat, "Quantifying scenic areas using crowd-sourced data and computer vision", *Royal Society Open Science*, vol. 2, no. 9, p. 150 460, 2015.

[61] T. Näscher, *Design and implementation of a lightweight decentralized federated learning platform*, University of Zurich, 2025.

[62] M. S. Tveit, Å. Ode, and G. Fry, "Key concepts in a framework for analysing visual landscape character", *Landscape Research*, vol. 31, no. 3, pp. 229–255, 2006.

[63] T. C. Daniel, "Whither scenic beauty? visual landscape quality assessment in the 21st century", *Landscape and Urban Planning*, vol. 54, no. 1, pp. 267–281, 2001.

Abbreviations

IAA	Image Aesthetic Assessment
FL	Federated Learning
DFL	Decentralized Federated Learning
CNN	Convolutional Neural Network
NIMA	Neural Image Assessment
AVA	Aesthetic Visual Analysis
SRCC	Spearman Rank Correlation Coefficient
PLCC	Pearson Linear Correlation Coefficient
IID	Independent and Identically Distributed
Non-IID	Non-Independent and Identically Distributed
DFL	Decentralized Federated Learning
FLyra	Decentralized Federated Learning Framework

List of Figures

List of Tables

Appendix A

Example Images from the Swiss Landscape Dataset

This appendix presents example landscape images collected for the Swiss Landscape Dataset. The images are collected from publicly available platforms, including Flickr and Wikimedia Commons. All images are used in accordance with their respective licenses for research and academic purposes. The examples illustrate the visual differences between low, medium, and high scenicness images.

Figure A.1: Example images with high scenicness scores.

Figure A.2: Example images with medium scenicness scores.

Figure A.3: Example images with low scenicness scores.