Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Master Thesis in ICT for Internet and Multimedia

Enhancing Point Cloud Quality Assessment with Grouped Convolutions: A Streamlined Approach Inspired by COPP-Net

Master Candidate

Asmaa Mirkhan

Student ID 2054616

Supervisor

Prof. Federica Battisti

University of Padua

Co-supervisor

Michael Neri

Rome Tre University

Academic Year
2023/2024

I dedicate this part of the thesis to thanking
my family, who serves as my super powerful, strong,
great, and precious army and to wholeheartedly
expressing my love for my brother, Amer,
who is the hero of my life.

Abstract

In the dynamic landscape of 3D vision applications, the Point Cloud Quality Assessment (PCQA) has become a critical focus. This paper presents an enhanced version of COPP-Net. COPP-Net strategically divides a point cloud into patches, leveraging a Point Cloud Pre-processing Module to normalize spatial coordinates and employ Farthest Point Sampling (FPS) and K-Nearest Neighbor algorithms for efficient patch creation. The subsequent Patch Feature Generation Module utilizes local texture and 3D structure feature generation Adaptive R-Sampling KNN PointNet++ Network networks ($ARKP_t$ and $ARKP_s$) based on the ARKP architecture. Notably, the Adaptive R-Sampling KNN PointNet++ Network (ARKP) network undergoes improvements, including grouped convolutions and block reduction, resulting in a remarkable 50% reduction in trainable parameters and enhanced computational efficiency. The Point Cloud Quality Regression Module predicts the overall point cloud quality score from patch features, employing a regression head with linear layers, batch normalization, and leaky ReLU layers. The Correlation Analysis Network (CORA) network further refines the assessment by estimating correlations between patch and overall point cloud quality, introducing correlation labels for improved accuracy. Experiments conducted on diverse datasets, including WPC, WPC2.0, and LS-PCQA, showcase the efficacy of the improved COPP-Net. Impressively, the introduced improvements result in a 20% decrease in one-epoch time for ARKP and a 10% decrease for CORA, while maintaining consistent model accuracy across all tested datasets.

Contents

List of Figures

List of Tables

List of Acronyms

AI Artificial Intelligence

ARKP Adaptive R-Sampling KNN PointNet++ Network

ASPRS American Society for Photogrammetry and Remote Sensing

BN Batch Normalization

CNN Convolutional Neural Network

CORA Correlation Analysis Network

DNN Deep Neural Network

FPS Farthest Point Sampling

GELU Gaussian Error Linear Units

ICP Iterative Closest Point

ISO International Organization for Standardization

IQA Image Quality Assessment

KNN K-Nearest Neighbor

LiDAR Light Detection and Ranging

MSE Mean Squared Error

MLP Multi-Layer Perceptron

NR-3DQA No-Reference 3 Dimensional Quality Assessment

NR-IQA No-Reference Image Quality Assessment

NR-PCQA No-Reference Point Cloud Quality Assessment

PCQA Point Cloud Quality Assessment

PLCC Pearson Linear Correlation Coefficient

ReLU Rectified Linear Unit

RMSE Root Mean Square Error

SfM Structure from Motion

SRCC Spearman Rank Correlation Coefficient

SSG Single-Scale Grouping

SSIM Structural Similarity

WPC Waterloo Point Cloud

1 INTRODUCTION

In the dynamic realm of 3D vision application, the assessment of point cloud quality plays a pivotal role, intricately influencing the precision and effectiveness of various technologies. From enhancing augmented reality experiences to guiding autonomous navigation systems, the importance of quality evaluation cannot be overstated [22].

Point clouds, which are elaborate spatial representations crafted from a multitude of data points, serve as the foundation for diverse tasks, ranging from precise object recognition [41] to the intricacies of spatial mapping [8]. However, the relentless pursuit of reliability and accuracy in these point clouds presents challenges that demand innovative solutions.

To comprehend the significance of our work, it is crucial to navigate through the complex landscape of Point Cloud Quality Assessment (PCQA) and understand its nuanced relevance. Point clouds, generated from 3D scanning devices or depth-sensing technologies, are indispensable representations of physical spaces [15]. Nevertheless, the inherent quality of point clouds faces challenges due to factors such as sensor noise, occlusions, or distortions, thereby introducing complexities to the PCQA task [41], the task that holds significance in various applications, including Immersive Technologies (e.g., VR, XR, MR), where the accurate assessment of point cloud quality becomes crucial.

In response, our research focuses on refining COPP-Net, an advanced No-Reference Point Cloud Quality Assessment (NR-PCQA) methodology introduced in [6]. NR-PCQA assesses the quality based on inherent features, statistical measures, or other characteristics present within the point cloud itself.

Accordingly, COPP-Net involves dividing a point cloud into patches, generating texture and structure features per patch, fusing them into patch features for quality prediction. Subsequently, it analyzes correlations among all patches to derive correlation weights. The final quality score is determined using the predicted patch qualities and correlation weights.

Within the broader context of PCQA, variations in quality within a point cloud pose formidable challenges. Existing methodologies may fall short by prioritizing global measures, overlooking subtle intricacies in specific regions. COPP-Net addressed this limitation by adopting a patch-based approach, recognizing the inherently local nature of quality variations—an essential step in enhancing the precision of PCQA.

Accordingly, our primary objective becomes clear; to enhance computational efficiency without significantly compromising the precision of quality assessment. The computational demands of processing large-scale point clouds guide our exploration towards optimization. Specifically, our focus is on recalibrating the ARKP and CORA structures within COPP-Net. The introduction of grouped convolutions, which was first introduced in AlexNet in 2012 [16], coupled with a strategic reduction in the number of blocks, aims to significantly reduce trainable parameters—an optimization journey balancing computational efficiency with the unwavering pursuit of accuracy.

Before delving into the details of our modifications, it is paramount to contextualize our work within the expansive landscape of PCQA. The upcoming chapters begin with the research background, providing fundamental concepts related to PCQA along with a detailed explanation of evaluation metrics used in this context. We then conclude by presenting related works already undertaken by other researchers in this field.

Subsequently, we delve into experimental analysis, explaining the fundamental Artificial Intelligence (AI) concepts used in the context of our work. Consequently, we comprehensively explain the structure of COPP-Net, highlighting its drawbacks, which led us to suggest improvements.

After that, we introduce our structural improvements over both networks presented in COPP-Net and report the results that we obtained, comparing them with the results acquired from the original network in terms of evaluation metrics, epoch time, and number of parameters.

Finally, we summarize our work, concluding our achievements and suggesting potential pathways for further exploration in the future.

2

RESEARCH BACKGROUND

In the vast and transformative landscape of 3D vision applications, it is essential to delve into the abstract concepts that form the foundation of PCQA. These concepts are the compass guiding the evaluation of the accuracy and fidelity of point clouds – three-dimensional representations that serve as digital reflections of physical spaces.

2.1 Point Cloud Quality Assessment - Fundamental Concepts

2.1.1 Introduction to Point Clouds

Point clouds represent a three-dimensional set of data points in space, capturing the surfaces and structures of physical objects or environments. Each point in a point cloud corresponds to a specific location in the scanned area, defined by its spatial coordinates and often associated with additional attributes such as color or intensity, as shown in the following figure 2.1. Point clouds are widely used in various fields, including remote sensing, geospatial mapping, computer vision, and augmented reality.

The generation of point clouds is typically achieved through advanced sensing technologies such as LiDAR (Light Detection and Ranging), photogrammetry, or structured light scanning. LiDAR, for instance, emits laser pulses and measures the time it takes for the laser beams to return after hitting objects in

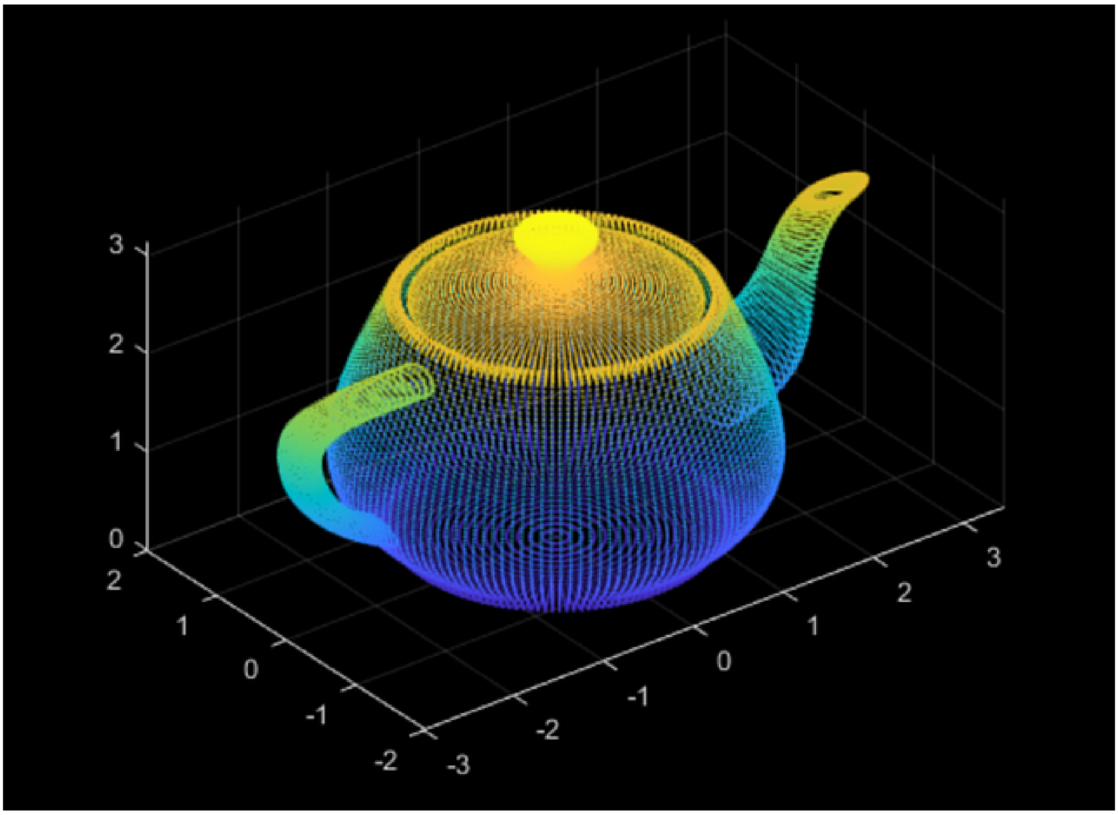


Figure 2.1: Point cloud example.

the environment. This process results in a dense and accurate representation of surfaces, creating a point cloud that faithfully captures the geometric details of the scanned area.

Point clouds play a pivotal role in creating realistic 3D models, conducting environmental assessments, and aiding in the development of autonomous systems. Understanding the fundamental concepts of point clouds, their acquisition methods, and inherent characteristics is crucial for effective PCQA, which involves evaluating the reliability and accuracy of the captured data. In the subsequent sections, we will explore the nuances of point cloud data, its characteristics, and the key considerations in assessing its quality.

2.1.2 Data Acquisition and Sources

Data acquisition is a crucial step in the creation of point clouds, involving the collection of spatial information to generate a comprehensive three-dimensional representation of an environment. Various methods are employed for acquiring point cloud data, each tailored to specific applications and scenarios.

Light Detection and Ranging (LiDAR) stands out as a primary technology for point cloud data acquisition. It utilizes laser beams to measure the distance to objects, generating precise spatial coordinates. Terrestrial LiDAR systems are ground-based and ideal for capturing detailed information in close-range environments, while airborne LiDAR is mounted on aircraft, providing efficient coverage for larger areas. Mobile LiDAR platforms, often integrated into vehicles and planes, offer flexibility in capturing dynamic environments such as urban landscapes, as shown in the figure 2.2.

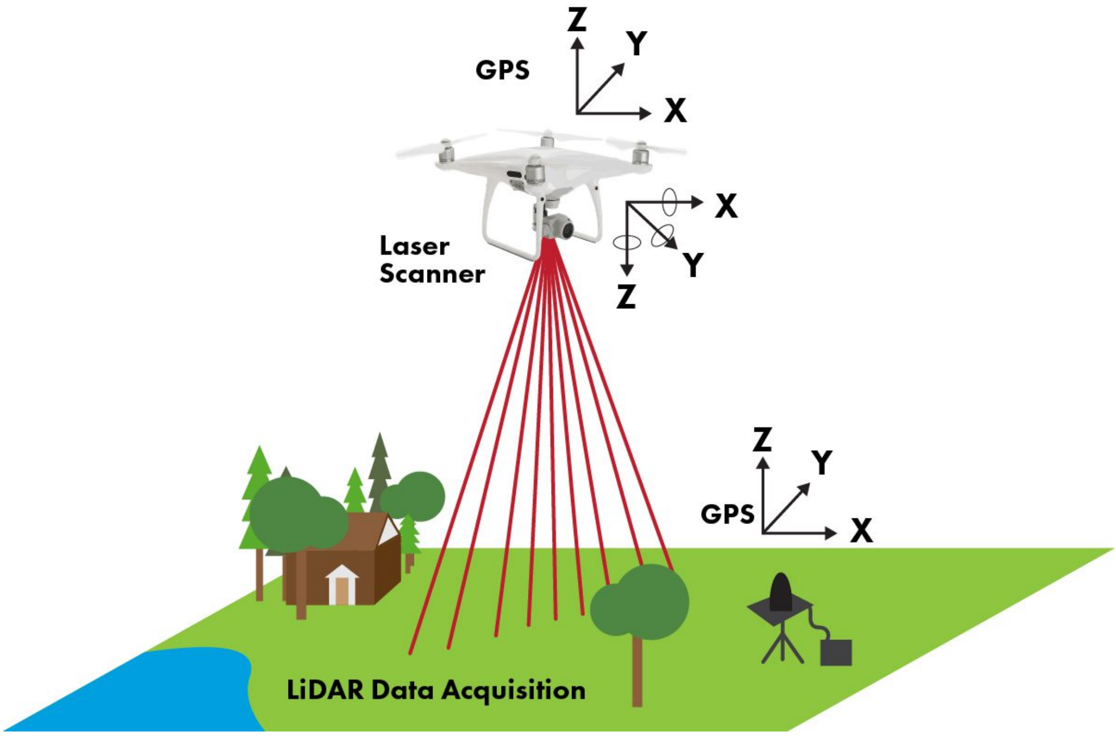


Figure 2.2: LiDAR Technology.

Photogrammetry is another widely used technique, relying on overlapping images captured from different viewpoints to derive 3D information. The Structure from Motion (SfM) algorithm extracts point clouds from these images, contributing to detailed and textured reconstructions [25].

Understanding the nuances of these data acquisition methods is crucial for evaluating the quality of resulting point clouds. In the subsequent sections, we will delve into the characteristics of point clouds generated through different sources, laying the foundation for effective PCQA.

2.1.3 Quality Metrics and Parameters

PCQA relies on a set of fundamental metrics and parameters that gauge the accuracy, completeness, and precision of the captured data. These metrics play a pivotal role in evaluating the reliability of point clouds for various applications.

- **Completeness:** Measures the percentage of the expected points that are successfully captured in the point cloud. Reflects the comprehensiveness of the data, highlighting any gaps or missing information.
- **Accuracy:** Assesses the closeness of the point cloud data to the true positions of the scanned objects. Influenced by the precision of the data acquisition system and the calibration accuracy of sensors.
- **Precision:** Involves the consistency of measurements within the point cloud. High precision indicates minimal variability in the spatial coordinates of points, enhancing the overall reliability of the data.
- **Density:** Measures the concentration of points in a given space, influencing the level of detail and fidelity in the representation. Higher point cloud density often leads to more accurate reconstructions.

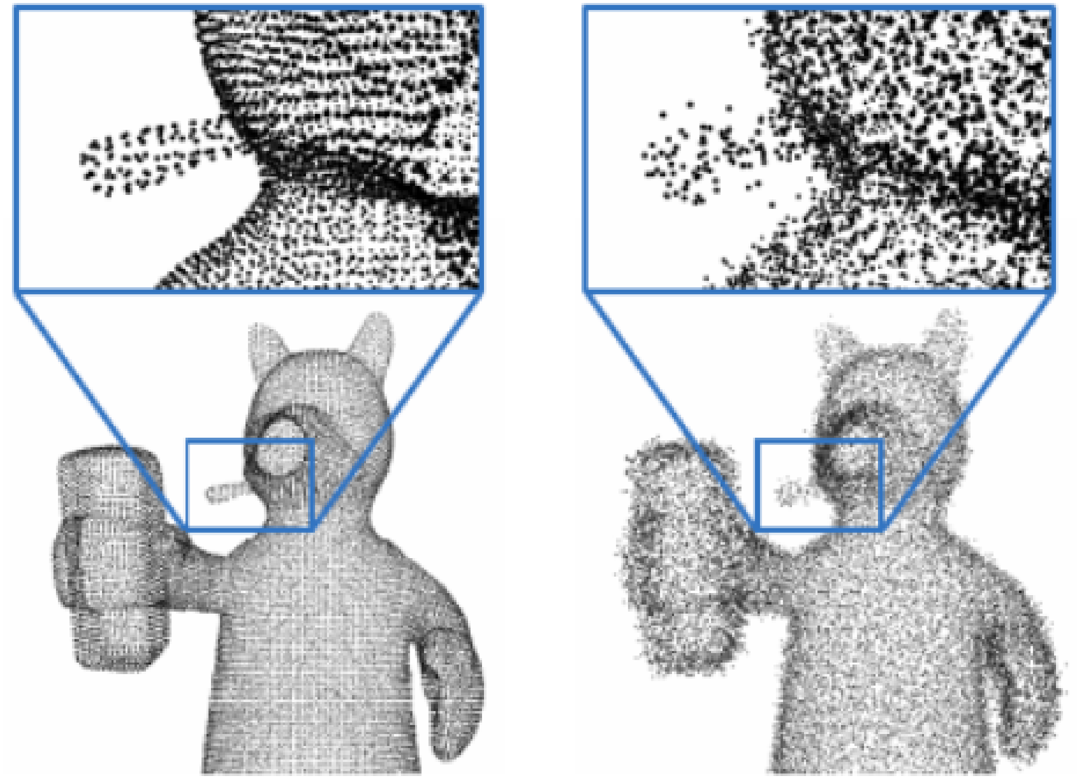
Understanding and quantifying these quality metrics are essential for researchers and practitioners to ensure that point clouds meet the desired standards for specific applications. In the subsequent sections, we will explore how these metrics, along with other parameters, contribute to a comprehensive PCQA framework.

2.1.4 Noise and Artifacts

Point cloud data, despite its precision and accuracy, is susceptible to various imperfections, often in the form of noise and artifacts. Understanding these nuisances is essential for comprehensive PCQA.

Noise in point clouds refers to random and unwanted variations in data, as shown in the figure 2.3 adapted from [10], typically arising from limitations in sensing technologies or environmental conditions. Common sources of noise include atmospheric interference in LiDAR systems or sensor inaccuracies. Evaluating and mitigating noise is crucial for ensuring the fidelity of the captured information.

Artifacts, on the other hand, are undesirable features introduced during the data acquisition process. They can result from factors such as occlusions, reflections, or inadequate calibration. Artifacts manifest as distortions or irregularities



(a) Ground truth. (b) Noisy.

Figure 2.3: Noise in point clouds.

in the point cloud, impacting its overall quality. Recognizing and addressing artifacts is vital for producing accurate and reliable 3D representations.

Several techniques are employed to mitigate noise and artifacts, ranging from advanced filtering algorithms to improved calibration procedures. Additionally, understanding the nature of noise and artifacts aids in the development of robust post-processing strategies.

In the subsequent sections, we will delve into specific types of noise and artifacts, their impact on point cloud quality, and methodologies to effectively identify and minimize their influence. A thorough exploration of noise and artifacts enhances the overall understanding of the challenges associated with point cloud data and contributes to the establishment of rigorous quality assessment frameworks.

2.1.5 Data Registration and Alignment

Data registration and alignment form a critical phase in the point cloud processing pipeline, influencing the accuracy and coherence of the final three-dimensional representation. This stage involves integrating multiple scans or datasets into a common coordinate system, ensuring seamless continuity and a unified perspective. The significance of precise registration and alignment cannot be overstated, particularly in scenarios where point clouds are generated from disparate sources or acquired at different times.

Registration, involves the spatial adjustment of individual point clouds to achieve a consistent reference frame. Various registration methods exist, including feature-based techniques that identify common features in different scans, and Iterative Closest Point (ICP) algorithms that iteratively refine the alignment based on point correspondences [4].

Alignment, closely related to registration, focuses on the transformation of point clouds to minimize discrepancies and optimize congruence. This process enhances the overall coherence of the combined datasets, facilitating accurate analyses and reconstructions.

Challenges in data registration and alignment arise from factors such as sensor inaccuracies, variations in data resolution, and environmental changes. Overcoming these challenges requires sophisticated algorithms capable of robustly handling complex scenarios. Efficient registration and alignment are essential for applications such as 3D modeling, virtual reality, and autonomous navigation. In the subsequent sections, we will explore advanced registration techniques, error mitigation strategies, and their pivotal role in ensuring the integrity of point cloud data for downstream analyses and applications.

2.1.6 Uncertainty and Error Modeling

In the realm of point cloud data, acknowledging and quantifying uncertainties and errors is paramount for achieving robust and reliable results. Uncertainty arises from various sources, including sensor limitations, environmental conditions, and the inherent complexity of capturing real-world geometry.

Uncertainty modeling involves characterizing the imprecision associated with each point in the cloud. This can stem from factors such as sensor noise, calibration errors, and atmospheric conditions during data acquisition. Developing models to represent these uncertainties aids in the creation of probabilistic frameworks, offering a more nuanced understanding of the reliability of the captured information.

Error modeling is closely tied to uncertainty and focuses on quantifying inaccuracies in point cloud data. This includes systematic errors introduced by sensor biases or calibration drift. Understanding error sources and patterns enables researchers and practitioners to implement corrective measures during data processing, improving the overall quality of the point cloud.

Common methods for uncertainty and error modeling include Monte Carlo simulations, where multiple virtual point cloud datasets are generated based on known uncertainties. Bayesian frameworks are also employed to express uncertainties probabilistically. Effectively addressing uncertainty and error in point cloud data enhances the credibility of subsequent analysis and applications. In the following sections, we will delve into specific techniques for uncertainty and error modeling, emphasizing their role in refining PCQA methodologies and promoting informed decision-making in various domains.

2.1.7 Data Filtering and Cleaning

The integrity and reliability of point cloud data heavily depend on the effectiveness of data filtering and cleaning processes. These processes are pivotal in removing unwanted noise, outliers, and artifacts, ensuring that the resulting point cloud accurately represents the scanned environment. In this section, we will explore various strategies and methodologies employed in data filtering and cleaning.

- **Noise Removal:** Noise, often introduced by sensor inaccuracies or environmental interference, can obscure the true structure of a scene. Statistical filtering techniques, such as mean or median filtering, are commonly applied to identify and eliminate random noise. These methods leverage statistical measures to distinguish between valid points and noise, contributing to a cleaner point cloud.

- **Outlier Detection and Removal:** Outliers, defined as points significantly deviating from the expected distribution, can adversely impact the accuracy of point cloud data. Robust algorithms, including spatial and morphological filters, are employed to detect and remove outliers. Iterative approaches, like the ICP algorithm, iteratively refine the alignment and filter out outliers, enhancing the overall data quality.

- **Artifact Mitigation:** Artifacts, unwanted features introduced during data acquisition, may include reflections, occlusions, or calibration errors. Advanced filtering methods, such as adaptive filtering and machine learning-based approaches, are employed to mitigate artifacts while preserving essential information. These techniques ensure a more accurate representation of the scanned scene.

- **Feature-Preserving Filtering:** While removing noise and outliers is crucial, preserving important features of the point cloud is equally important. Feature-preserving filters, such as bilateral or edge-preserving filters, aim to retain edges and structural details while reducing noise. These filters are particularly beneficial when maintaining fine details is essential for downstream applications.

- **Machine Learning Applications:** Recent advancements in machine learning have brought forth novel approaches for data filtering and cleaning. Supervised learning models can be trained to classify points as valid or outliers based on learned patterns. Neural networks, particularly CNNs, demonstrate effectiveness in identifying complex patterns and distinguishing between noise and valid data.

The selection of filtering and cleaning strategies depends on the characteristics of the point cloud and the specific goals of the application. Balancing the removal of unwanted elements with the preservation of critical information is a nuanced process. In the subsequent sections, we will delve into specific algorithms, case studies, and best practices for data filtering and cleaning, providing a comprehensive understanding of these critical steps in PCQA.

2.1.8 Validation and Ground Truth

Ensuring the accuracy and reliability of point cloud data necessitates rigorous validation against ground truth information. Validation serves as a crucial step in PCQA, providing a means to assess the performance of data acquisition systems, processing algorithms, and overall data fidelity.

- **Ground Truth Establishment:** Establishing ground truth involves acquiring reference data with known accuracy to compare against the generated point cloud. This can be achieved through traditional surveying methods, high-precision instruments, or other reliable sources depending on the application. Ground truth data provides a benchmark for evaluating the quality of the point cloud and identifying potential errors or discrepancies.

- **Validation Metrics:** Various metrics are employed to quantitatively assess the alignment and accuracy of the point cloud concerning ground truth. Common metrics include Root Mean Square Error (RMSE), which measures the average difference between corresponding points in the point cloud and ground truth [5]. Other metrics include precision, recall, and F1 score, providing comprehensive insights into the performance of the data.

- **Comparative Analysis:** Comparing the point cloud against ground truth involves spatially aligning the two datasets and analyzing their congruence. Visualization tools and color mapping techniques can aid in highlighting areas of agreement and discrepancies. Comparative analysis not only validates the accuracy of the point cloud but also identifies potential areas for improvement in data acquisition or processing methodologies.

- **Dynamic Environments:** In dynamic environments where the scene may change over time, continuous validation is crucial. This involves periodic updates of ground truth data to reflect changes in the environment and ensure the ongoing accuracy of the point cloud. Real-time validation mechanisms may also be employed in dynamic scenarios.

- **Challenges and Considerations:** Challenges in validation include the need for high-precision ground truth data, potential biases in the selection of validation metrics, and addressing dynamic or evolving scenes. Researchers must carefully consider these challenges to establish a robust validation framework.

In the subsequent sections, we will delve into specific case studies, methodologies, and advancements in the validation of point cloud data. By exploring the intricacies of validation and ground truth, we aim to provide a comprehensive understanding of the reliability and accuracy of point cloud datasets in diverse applications.

2.1.9 Standards and Guidelines

In the rapidly evolving field of point cloud technology, the establishment of standards and guidelines is instrumental for ensuring consistency, interoperability, and quality across diverse applications. Standardization efforts are essential to harmonize data acquisition, processing, and representation methodologies, fostering a common framework for practitioners and researchers.

International Standards: Organizations such as the International Organization for Standardization (ISO) and the American Society for Photogrammetry and Remote Sensing (ASPRS) have played pivotal roles in developing and maintaining international standards for point cloud data. ISO 19115-1, for example, outlines standards for metadata, while ISO 19107 defines the spatial schema for geographic information. ASPRS standards cover LiDAR data exchange formats and classification specifications, ensuring a standardized approach to data representation.

Data Exchange Formats: Standardized data exchange formats facilitate seamless interoperability between different software and hardware systems. The LAS (LiDAR Data Exchange Format) and ASTM E57 are widely adopted standards for exchanging point cloud data. These formats ensure that data captured by one system can be effectively utilized by another, promoting collaboration and data sharing.

Quality Assessment Guidelines: Guidelines for PCQA are essential for practitioners to conduct robust evaluations of data quality. These guidelines often include recommended metrics, procedures for ground truth establishment, and best practices for noise reduction and outlier removal. Adhering to these guidelines ensures a standardized and systematic approach to assessing point cloud quality.

Emerging Challenges and Future Directions: As technology advances and applications diversify, the establishment of new standards becomes imperative. Addressing emerging challenges such as dynamic scene capturing, point cloud compression, and semantic labeling requires ongoing efforts in standardization. Collaborative initiatives between industry, academia, and standards organizations are essential to adapting standards to the evolving landscape of point cloud technology.

Implementation in Industries: The adoption of standards is particularly crucial in industries such as geospatial mapping, urban planning, and autonomous navigation. Compliance with established standards enhances data reliability, facilitates data integration, and supports the development of interoperable solutions across various sectors.

2.2 PCQA - Metrics

In the domain of PCQA, several metrics are commonly used to evaluate the performance of algorithms in comparison to human perception or ground truth quality assessments. In this section, we will discuss three key metrics: Spearman Rank Correlation Coefficient (SRCC), Pearson Linear Correlation Coefficient (PLCC), and RMSE.

2.2.1 SRCC

SRCC is a non-parametric measure that assesses the monotonic relationship between two variables. In the context of PCQA, SRCC is often used to evaluate the agreement between the rankings of quality scores assigned by human observers (X) and those predicted by an algorithm (Y). The formula for SRCC is given by:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{2.1}$$

where d_i is the difference between the ranks of corresponding pairs of observations and n is the number of observations.

2.2.2 PLCC

PLCC measures the linear relationship between two variables. In PCQA, PLCC is employed to assess the correlation between quality scores assigned by human observers (X) and those predicted by an algorithm (Y). The formula for PLCC is given by:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}} \tag{2.2}$$

where X_i and Y_i are the individual observations, and $\bar{X}$ and $\bar{Y}$ are the means of X and Y, respectively.

2.2.3 RMSE

Root Mean Square Error is a metric that quantifies the average magnitude of errors between predicted values and observed values. In PCQA, RMSE can be used to measure the accuracy of an algorithm's predictions compared to ground truth. The formula for RMSE is given by:

$$RMSE = \sqrt{\frac{\sum(Y_i - X_i)^2}{n}} \tag{2.3}$$

where X_i and Y_i are the observed and predicted values, respectively, and n is the number of observations.

These metrics provide valuable insights into the performance of PCQA algorithms and their alignment with human perception.

2.3 PCQA - Approaches

In this section, we explore different approaches to PCQA. Before delving into specific experiments, let's provide a brief overview of both Full-Reference and No-Reference approaches that already done by researchers, setting the stage for the methods that will be discussed in the following sections.

2.3.1 Full-Reference Metrics

Full-Reference (FR) metrics in the domain of point cloud quality assessment (PCQA) have played a pivotal role, initially developed to evaluate the efficacy of point cloud compression methods. Noteworthy among these classic metrics are `p2point` [23] and `p2plane` [32], both proposed by the Moving Picture Experts Group (MPEG). Despite their advantageous low computational complexity, these metrics exhibited limitations in accuracy, particularly when confronted with complex distortion types, leading to potential instability in assessment results.

Recognizing these limitations, researchers delved into alternative approaches to enhance the precision of FR metrics. Alexiou et al. [2] introduced a straightforward yet efficient objective metric designed to capture perceptual degradations in distorted point clouds. This metric demonstrated superiority in predicting visual quality under realistic distortions, such as octree-based compression, establishing its excellence through correlation with subjective quality assessment scores.

Meynet et al. [24] contributed to the field with PC-MSDM, a full-reference metric grounded in local curvature statistics. An extension of the MSDM metric tailored for 3D meshes, PC-MSDM showcased its capabilities when evaluated on an open subjective dataset of point clouds compressed by octree pruning. The results indicated that PC-MSDM outperformed its counterparts, exhibiting a higher correlation with mean opinion scores.

Addressing the challenges posed by geometric point cloud distortions, Javaheri et al. [14] proposed a generalized Hausdorff distance-based quality metric. By varying the generalized Hausdorff distance parameter, the authors derived

15 different distances, offering a unique perspective on quality metrics. This innovative approach not only outperformed state-of-the-art MPEG PC geometry quality metrics but also introduced a novel framework for assessing point cloud quality.

Drawing on the principles of the Structural Similarity (SSIM) Index, Wang et al. [33] constructed an SSIM quality measure tailored for point clouds. The authors explored structural information in point clouds by effectively separating the influence of illumination. This unique perspective added depth to full-reference metrics in the context of PCQA.

Yang et al. [35] introduced GraphSIM, a metric designed to accurately and quantitatively predict human perception of point clouds with superimposed geometry and color impairments. Motivated by the sensitivity of the human vision system to high spatial-frequency components, GraphSIM utilized graph signal gradient as a quality index. This metric demonstrated state-of-the-art performance across various impairments, providing notable gains in predicting subjective mean opinion scores (MOS).

In a comprehensive study, Liu et al. [20] built a large 3D point cloud database for subjective and objective quality assessment. They proposed a novel objective PCQA model based on the principle of information content-weighted structural similarity. The model not only correlated well with subjective opinions but also significantly outperformed existing PCQA models. This contribution underscored the importance of considering information content in structural similarity metrics.

These diverse contributions exemplify the evolving landscape of Full-Reference metrics in PCQA, incorporating insights from perceptual degradation capture, local curvature statistics, generalized distance-based metrics, and information content-weighted structural similarity. The pursuit of improved accuracy and robustness remains a driving force in this dynamic field of research, where each advancement adds a unique layer to our understanding of PCQA.

2.3.2 No-Reference Metrics

The field of No-Reference (NR) metrics in NR-PCQA has undergone significant exploration, driven by the imperative to surmount limitations posed by the scarcity of point cloud quality databases. As researchers sought methods capable of operating without access to reference data, innovative NR-PCQA strategies began to emerge, leveraging advancements in No-Reference Image Quality Assessment (NR-IQA) as a foundational framework. The maturation of NR-IQA methods, as exemplified by Liu et al. [21], provided the impetus for researchers to explore avenues for projecting point clouds into 2D representations, thereby enabling the utilization of traditional Image Quality Assessment (IQA) methods or CNNs for indirect quality prediction.

In the realm of No-Reference 3 Dimensional Quality Assessment (NR-3DQA), Zhang et al. [40] presented a comprehensive NR quality assessment metric for colored 3D models, embracing both point clouds and meshes. The method involved the projection of 3D models into quality-related geometry and color feature domains, utilizing 3D natural scene statistics (3D-NSS) and entropy to extract quality-aware features. Subsequently, machine learning techniques were employed to regress these features into visual quality scores. The proposed method underwent validation on the colored PCQA database (SJTU-PCQA), the Waterloo Point Cloud (WPC) dataset, and the colored mesh quality assessment database (CMDM). The results demonstrated the superiority of the proposed method, outperforming most compared NR 3D-QA metrics with competitive computational resources and significantly narrowing the performance gap with state-of-the-art Full-Reference 3D-QA metrics.

Yang et al. [36] introduced IT-PCQA, a novel NR-PCQA metric for 3D point clouds. Recognizing the compelling performance of deep neural networks in no-reference metric design, they addressed one of the most challenging issues in NR-PCQA – the lack of large-scale subjective databases to drive robust networks. Leveraging the rich subjective scores of natural images, the authors formulated an approach to quest the evaluation criteria of human perception via a DNN and transfer the capability of prediction to 3D point clouds. The proposed method, utilizing unsupervised adversarial domain adaptation, treated natural images as the source domain and point clouds as the target domain. They introduced a hierarchical feature encoder and a conditional discriminative network to extract effective latent features and minimize domain discrepancies. Experimental results showcased the proposed method's higher performance compared to traditional No-Reference metrics, even yielding results comparable to Full-Reference metrics. This approach not only demonstrated the feasibility of assessing the quality of specific media content without expensive and cumbersome subjective evaluations but also showcased the potential of deep learning in addressing the challenges of NR-PCQA.

In the context of NR-PCQA, Liu et al. [22] addressed the limitations of existing NR-PCQA research, primarily due to the lack of large-scale PCQA datasets. They took a significant step by constructing the large-scale PCQA dataset named LS-PCQA, consisting of 104 reference point clouds and more than $22,000$ distorted samples. Each reference point cloud in the dataset was augmented with 31 types of impairments at seven distortion levels. The authors proposed ResSCNN, a NR metric based on sparse CNNs, to accurately estimate the subjective quality of point clouds. The experimental results demonstrated that ResSCNN exhibited state-of-the-art (SOTA) performance among all existing NR-PCQA metrics and even outperformed some Full-Reference metrics.

Chetouani et al. [7] delved into the realm of blind (no-reference) scenarios in PCQA, where the original point cloud is unavailable. They proposed a two-step procedure, initially extracting three relevant low-level features from 10-cal patches, including geometric distance, local curvature, and luminance values. Subsequently, they employed a deep neural network to learn, from these low-level features, a mapping to the PC ground truth mean opinion score. Despite the limited availability of subjectively annotated datasets, the proposed approach exhibited potential on two state-of-the-art PC quality datasets, showcasing the viability of learning features from data using deep neural networks in a no-reference scenario.

This comprehensive exploration of No-Reference metrics in PCQA underscores the growing interest in methods that can operate without reliance on reference data. As databases for point cloud quality continue to expand, these sophisticated and robust methods are poised to play a pivotal role in addressing the challenges posed by the scarcity of reference information.

3

EXPERIMENTS AND ANALYSIS

In the pursuit of advancing PCQA, our analysis and experiments revolve around refining the COPP-Net, a NR-PCQA method at the forefront of 3D vision applications. The essence of our investigation lies in the optimization of COPP-Net, addressing the computational complexities associated with large-scale point cloud datasets. Motivated by the evolving demands of 3D vision applications, we embark on a journey to enhance the efficiency and effectiveness of COPP-Net, with a particular focus on its ARKP and CORA networks. Our modifications, including the integration of grouped convolutions in ARKP and selective block removal in CORA, aim to strike a balance between computational efficiency and model accuracy.

This phase of our research delves into the outcomes of these strategic adjustments. Through rigorous experimentation and analysis, we explore the impact of reduced trainable parameters on computational efficiency without compromising the precision of CORA. Our objective is to provide insights that extend beyond theoretical enhancements, emphasizing practical implications for real-world applications. The subsequent sections detail the methodology employed, present quantitative results across diverse datasets, and engage in a comprehensive discussion that interprets our findings and positions them within the broader landscape of CORA research.

3.1 AI in PCQA - Fundamental Concepts

AI has witnessed remarkable advancements, revolutionizing various domains. This section provides an in-depth exploration of foundational theoretical concepts crucial for understanding the design and functionality of neural network models. These concepts play a pivotal role in the context of Point Cloud Quality Assessment.

3.1.1 Convolutional Neural Networks (CNNs)

CNNs serve as the cornerstone of computer vision, excelling in extracting hierarchical features from structured grid data. Their convolutional layers employ filters to scan input data, capturing patterns like edges, textures, and intricate structures, as depicted in Fig 3.1 adapted from [3]. These layers are complemented by pooling operations, which reduce spatial dimensions while retaining critical features, enabling CNNs to comprehend complex relationships within images.

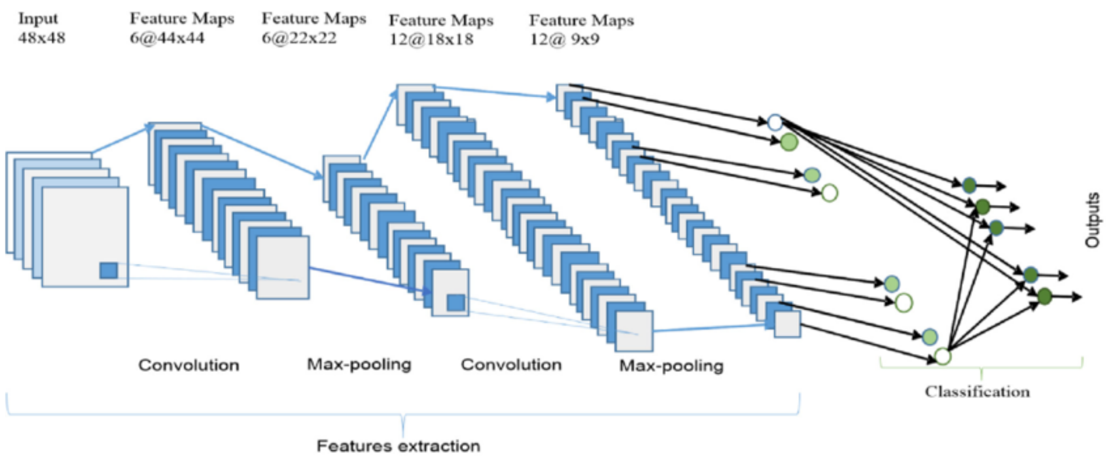


Figure 3.1: CNN example.

Convolutional layers play a crucial role by applying filters to input data, capturing local patterns, and contributing to the network's hierarchical learning capabilities, as illustrated in Fig 3.2 adapted from [28].

Grouped convolutions were first introduced in Alexnet [16] in 2012, they refine standard convolutional operations. In a typical convolutional layer, filters convolve across all input channels, which can be computationally demanding. Grouped convolutions introduce the concept of groups, dividing input channels

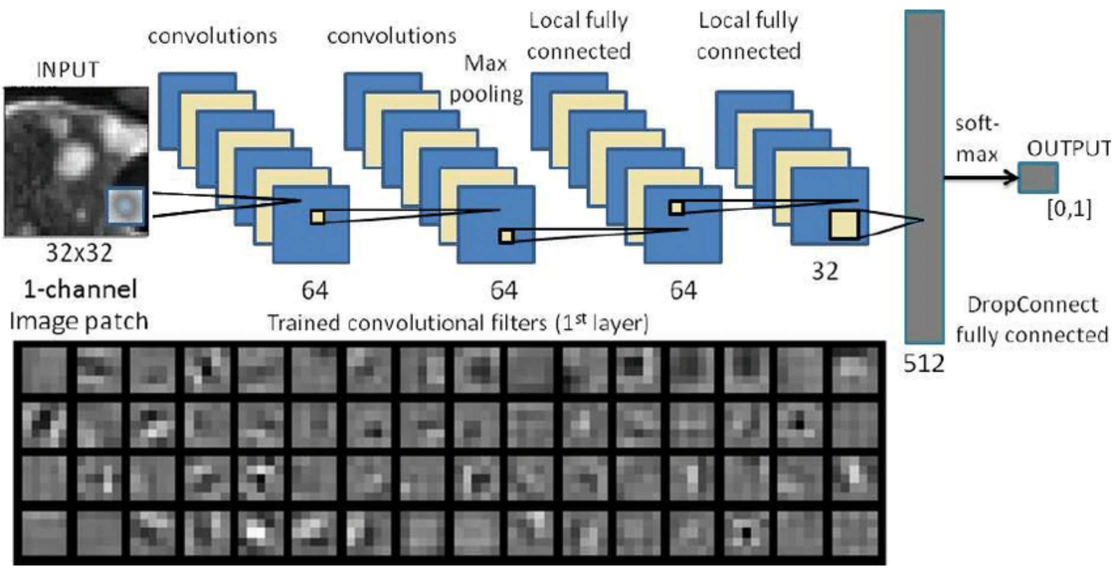


Figure 3.2: CNN layers example.

into subsets, and each group is convolved independently, as seen in Fig 3.3 adapted from [38].

Mathematically, grouped convolutions can be represented as:

$$\text{Output}_i = \text{Conv}(\text{Input}_i, \text{Filter}_i) \quad \text{for each group } i \tag{3.1}$$

Key advantages include parameter efficiency, parallelization, and adaptability to diverse computational resources. The utilization of grouped convolutions in CNN architectures enhances computational efficiency without compromising the ability to capture intricate patterns.

Pooling layers, often integrated after convolutional layers, facilitate spatial dimension reduction. Max pooling and average pooling, commonly used, preserve essential features while downsampling the data, enhancing computational efficiency, as visualized in Fig 3.4 adapted from [37].

3.1.2 Batch Normalization

The training of Deep Neural Network (DNN)s introduces challenges related to internal covariate shift, where the distribution of internal activations fluctuates during training. Addressing this issue, Batch Normalization (BN) normalizes layer inputs across mini-batches [13]. This normalization stabilizes and expedites training, alleviating problems associated with vanishing or exploding

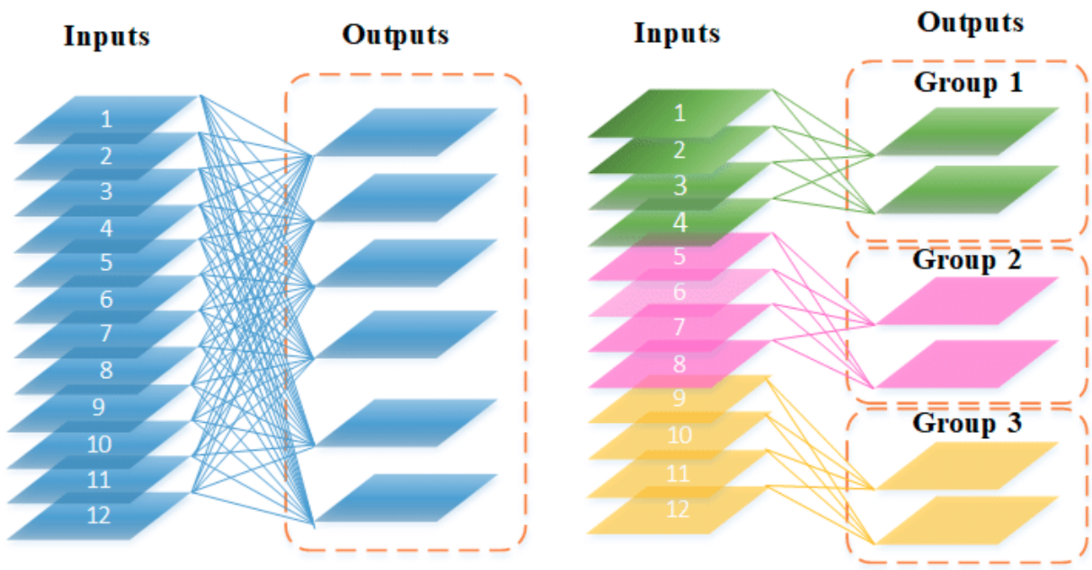


Figure 3.3: Comparison between standard convolutions and grouped convolutions.

gradients.

Internal covariate shift, the change in the distribution of internal activations during training [13], can hinder convergence and impact optimization effectiveness. BN effectively counters internal covariate shift by maintaining stable activations, facilitating more efficient learning.

BN's primary role is to stabilize training by normalizing inputs, thereby reducing internal fluctuations. This normalization contributes to a consistent learning rate, preventing saturation or divergence issues. The stabilizing effect of BN proves crucial for training DNNs effectively.

3.1.3 Activation Functions in Neural Networks

Activation functions are pivotal in infusing non-linearity into neural networks, allowing them to model intricate relationships. Among the array of choices, Rectified Linear Unit (ReLU) stands out as a widely embraced activation function recognized for its simplicity and effectiveness in mitigating the vanishing gradient problem.

The mathematical representation of ReLU is given by:

$$f(x) = \max(0, x) \tag{3.2}$$

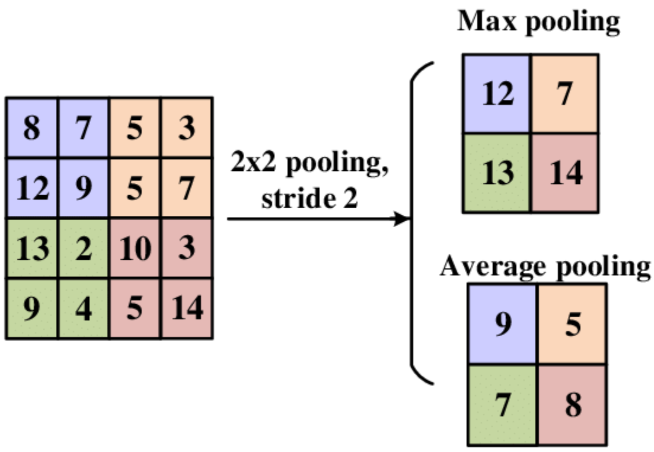


Figure 3.4: Pooling layers.

This formula clearly indicates that positive inputs yield the same output, while negative inputs result in zero output, as illustrated in Fig 3.5 adapted from [29].

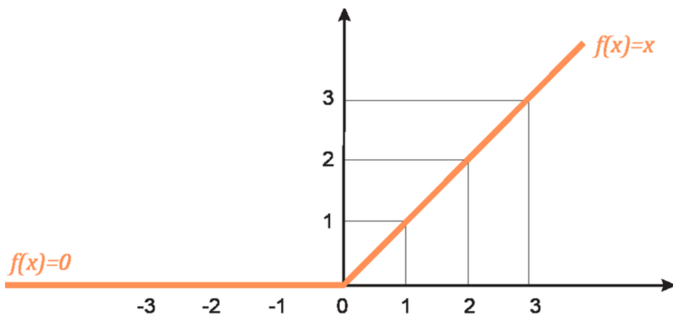


Figure 3.5: ReLU graph.

To counter the "dying ReLU" problem, a variant known as Leaky ReLU introduces a small, non-zero gradient for negative inputs. This adjustment ensures that neurons remain active during training, addressing one of ReLU's limitations.

The dying ReLU problem arises when neurons consistently output zero during training due to consistently negative inputs. Leaky ReLU mitigates this issue by allowing a small gradient for negative inputs, promoting the flow of information and preventing neurons from becoming inactive.

Beyond its role in preventing the "dying ReLU" problem, ReLU introduces non-linearity to the network, enhancing its capacity to model complex relationships. Its simplicity and effectiveness make ReLU a preferred choice in various neural network architectures. Leaky ReLU further bolsters the network's stability by addressing the "dying ReLU" problem.

As a viable option, Leaky ReLU emerges as an activation function, addressing some limitations of traditional ReLU [34]. In contrast to ReLU, Leaky ReLU introduces a small slope for negative input values, as depicted in Fig 3.6. This "leak" is mathematically expressed as:

$$f(x) = \max(\alpha x, x) \tag{3.3}$$

where α is a small positive constant, typically set to 0.01. The inclusion of this slope mitigates the "dying ReLU" problem, enhancing the robustness of neural networks.

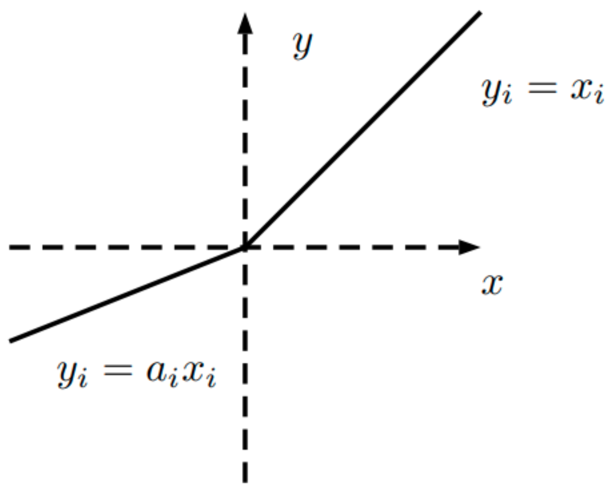


Figure 3.6: Leaky ReLU graph.

Alternatively, the GELU serves as an activation function offering a smooth approximation of the Rectified ReLU [9]. Defined by:

$$GELU(x) = \frac{1}{2}x\left(1 + \tanh\left(\sqrt{\frac{2}{\pi}}\left(x + 0.044715x^3\right)\right)\right) \tag{3.4}$$

GELU introduces non-linearity while maintaining smoothness, See its graph in 3.7 adapted from [17]. This characteristic can be advantageous in the optimization process during deep neural network training, addressing some limitations of traditional ReLU activations.

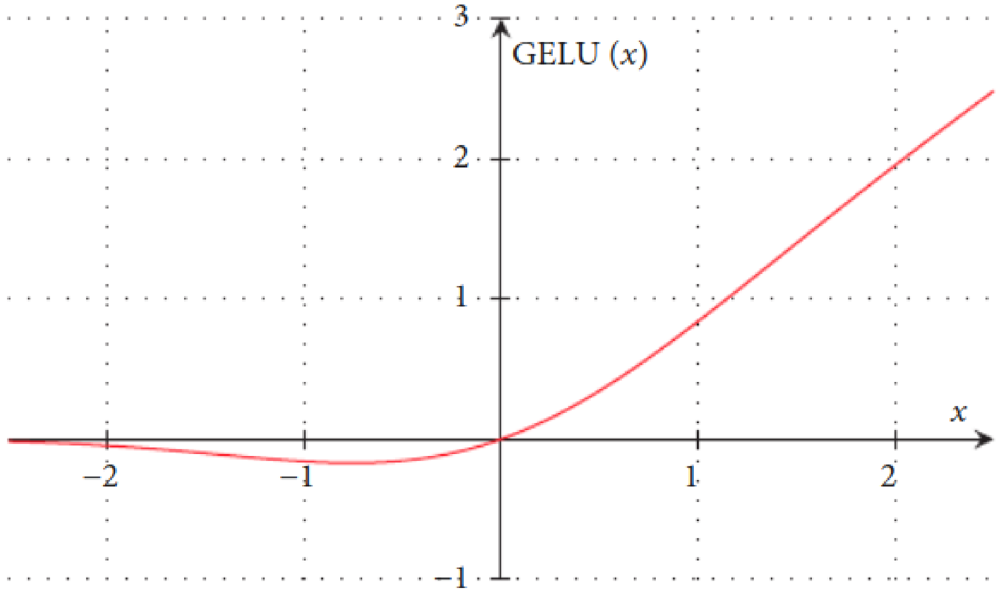


Figure 3.7: Leaky GELU graph.

3.1.4 Adaptive Pooling

Traditional pooling operations, relying on fixed-size regions, pose limitations in adapting to varying input sizes. In contrast, adaptive pooling, as exemplified by `F.adaptive_max_pool1d`, dynamically adjusts to input sizes, providing flexibility in handling variable-sized data. Particularly beneficial in the context of point clouds, adaptive pooling effectively addresses the irregular and unstructured nature of the data.

Adaptive pooling proves advantageous when dealing with variable-sized data, such as point clouds [11]. Its ability to adapt to different input sizes ensures the preservation of crucial information during spatial dimension reduction, proving crucial for effective point cloud processing.

Beyond handling variable-sized data, the adaptability introduced by adaptive pooling extends to enhancing the flexibility of neural network architectures. This flexibility accommodates variations in input sizes, making it well-suited for tasks where the size of the input data may vary. Consequently, models can effectively handle diverse datasets with varying spatial characteristics, demonstrating the adaptability of adaptive pooling in a broader architectural context.

3.1.5 Dropout

Dropout, a regularization technique designed to prevent overfitting during training, operates by randomly setting a fraction of input units to zero during each update. This process, illustrated in Fig 3.8 adapted from [27], effectively "drops out" some neurons, promoting robust learning by preventing the network from overly relying on specific features.

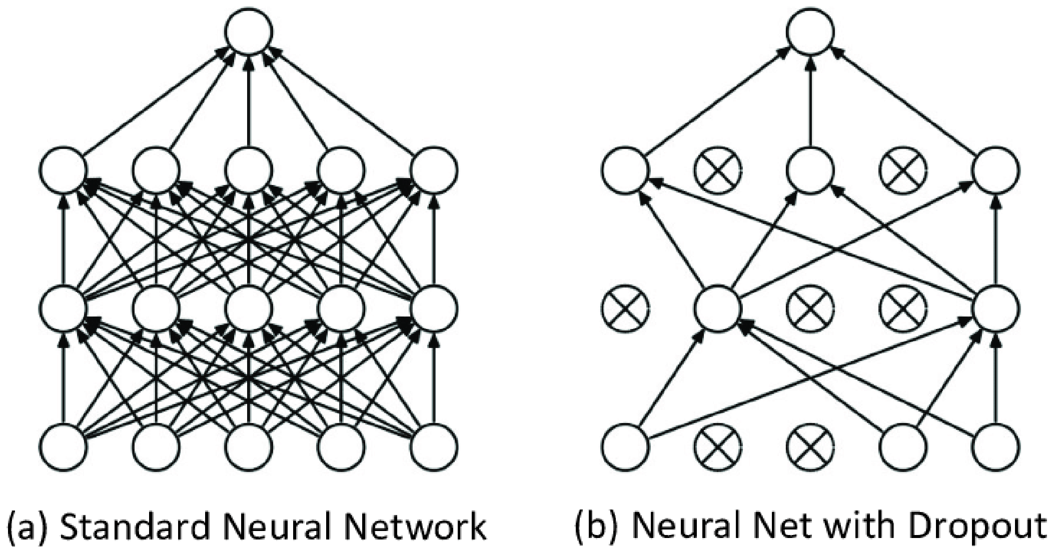


Figure 3.8: Dropout Example.

In addition to preventing overfitting, dropout introduces randomness during training, compelling the network to adapt to different subsets of features. This stochastic process contributes significantly to improved model generalization by preventing the learning of noise in the training data. Furthermore, dropout's stochastic nature brings about an ensembling effect during training, simulating the training of multiple models with different subsets of neurons active. This effect enhances the model's ability to generalize well to unseen data, encouraging the network to learn more robust features representative of the underlying data distribution. The ensembling effect adds an extra layer of adaptability to the network's learning process.

While dropout is a powerful technique, it is crucial to acknowledge that it introduces a degree of training instability. The stochastic nature of dropout may result in increased variability during training, and careful consideration is necessary to balance its advantages with potential drawbacks in specific scenarios.

3.1.6 Point Cloud Processing

The unique challenges posed by point clouds, characterized by their irregular and unstructured nature, necessitate specialized techniques for effective neural network processing. In contrast to grid-based data, point clouds lack a predefined order among their points, demanding tailored approaches. Point cloud processing involves a comprehensive set of methods encompassing feature extraction, noise reduction, and spatial alignment, each designed to address these distinctive characteristics.

The irregular and unstructured nature of point clouds, stemming from the absence of a predefined order among points, challenges conventional neural network architectures optimized for grid-based data. Consequently, the development of specialized processing techniques becomes imperative to handle point clouds effectively, taking into account their unique spatial arrangements.

Feature extraction stands out as a critical step in deriving meaningful insights from raw point data within point clouds. This involves employing various methods, such as local operations, neighborhood sampling, and adaptive pooling, to extract features that represent the underlying structures within the point cloud. The transformation of raw data into a format conducive to subsequent analysis is a primary objective of the feature extraction process.

Noise reduction is paramount for enhancing the quality of information within point clouds, given the presence of noise resulting from sensor inaccuracies or environmental factors. Techniques like filtering and outlier removal play a crucial role in ensuring robust feature extraction, thereby contributing to improved performance in subsequent tasks.

Spatial alignment emerges as another integral aspect of point cloud processing, enabling the comparison and analysis of different scans. Methods like ICP algorithms ensure precise registration of point clouds in a common coordinate system [12]. This precision in spatial alignment fosters meaningful comparisons and assessments, allowing the model to discern spatial relationships between different parts of the point cloud.

3.1.7 Self-Attention Mechanism

The self-attention mechanism, a foundational concept in neural networks, empowers models to evaluate the significance of different elements within a sequence, as depicted in Fig 3.9 adapted from [39]. Leveraging the Masked Causal Attention module, this mechanism proves instrumental in capturing long-range dependencies within input sequences, notably enhancing the model's capacity to discern relationships in point cloud data.

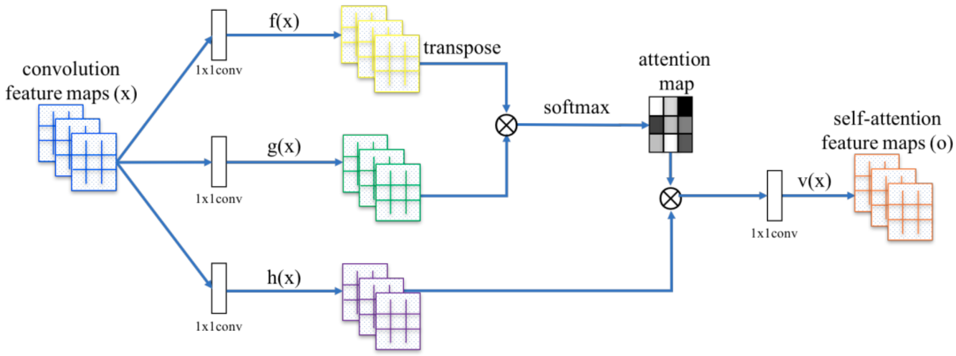


Figure 3.9: Self-Attention mechanism.

Beyond its fundamental application, self-attention allows the model to capture dependencies between distant elements in a sequence. In the realm of point cloud processing, this capability proves invaluable for understanding spatial relationships between points, even when they are far apart. The Masked Causal Attention module employs self-attention to enrich the model's contextual understanding, enabling it to capture intricate relationships within point cloud data, as illustrated in Fig 3.10 adapted from [30].

3.1.8 Network Optimization

At the heart of our methodology lies the strategic optimization of the ARKP and CORA networks embedded within the COPP-Net architecture. The ARKP network undergoes a pivotal transformation with the introduction of grouped convolutions. This meticulous enhancement is not merely a reduction in the number of trainable parameters; it is a thoughtful refinement aimed at addressing computational overhead without compromising the model's ability to discern intricate details in point cloud data. Simultaneously, the CORA network experiences a renaissance through selective block removal, streamlining

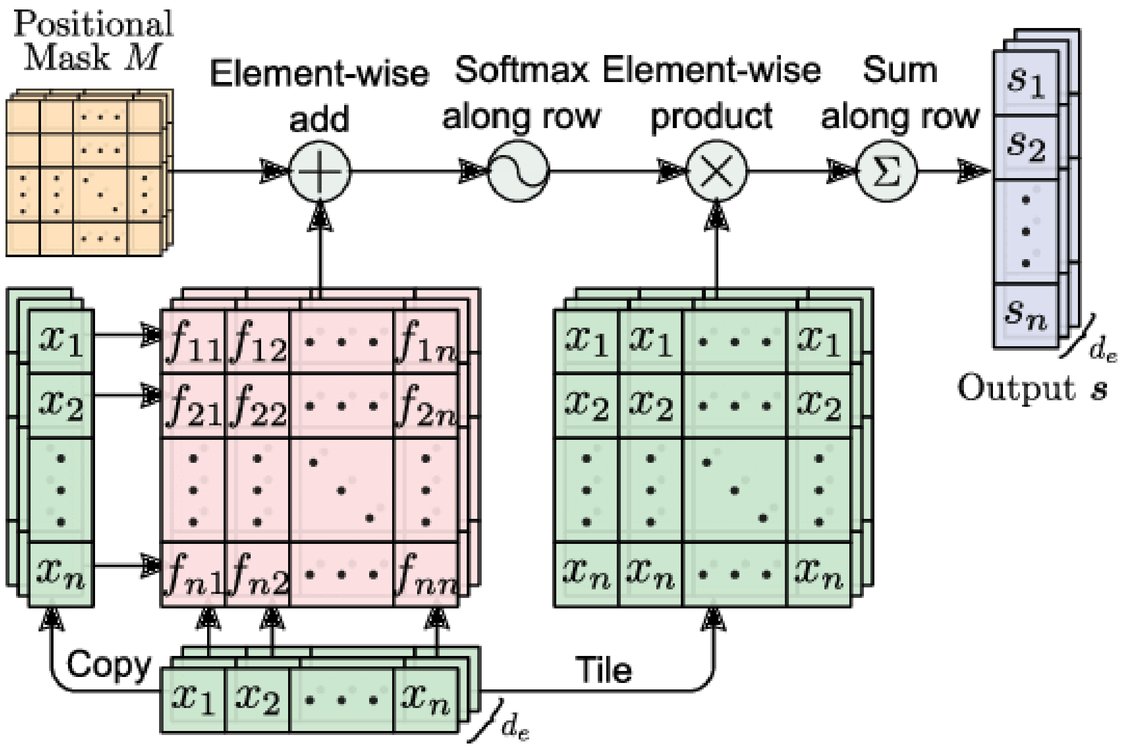


Figure 3.10: Masked Self-Attention mechanism.

its functionality for improved efficiency in estimating correlations between patch quality and overall point cloud quality. This phase represents a meticulous fine-tuning process to ensure that the COPP-Net is not just optimized but tailored to meet the demands of diverse point cloud datasets.

3.2 Approach of COPP-Net

3.2.1 Point Cloud Preprocessing Module

COPP-Net [6] recognizes the challenge posed by local quality variance within a point cloud and responds with a carefully designed preprocessing module. The module's primary goal is to address the inherent diversity in quality across different areas of a point cloud. To achieve this, the point cloud is divided into patches, considering that different parts of the point cloud may exhibit different quality correlations. The preprocessing begins with the normalization of spatial coordinates and color information. The spatial coordinates (x, y, z) are normalized into a sphere with a radius of 1000. Farthest Point Sampling (FPS) [26] is then employed to obtain a set of center points (C), and the K-Nearest

Neighbor (KNN) algorithm [1] is used to sample the nearest K points to each center point, forming the patches. This careful selection of patches ensures that the subsequent analysis takes into account the nuances of local quality distribution within the point cloud. Look Fig 3.11 for detailed structure of COPP-Net.

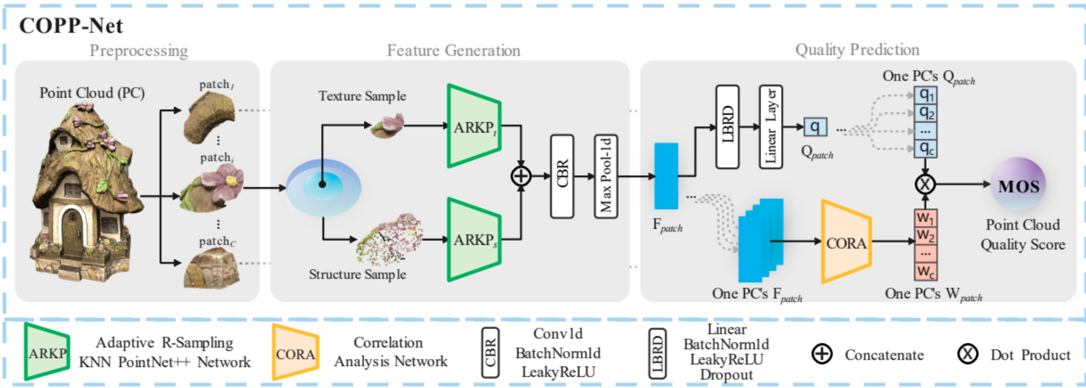


Figure 3.11: Detailed structure of COPP-Net [6].

3.2.2 Patch Feature Generation Module

The feature generation module within COPP-Net is a critical component responsible for generating informative features for texture and structure. Understanding the inherent differences in sensitivity to downsampling between texture and structure information, COPP-Net employs distinct strategies for generating these features.

Texture information, being sensitive to downsampling, is derived from continuous regions with similar quality scores. This is accomplished through the local texture feature generation network (ARKPt) and the 3D structure feature generation network (ARKPs). Both networks are based on the ARKP architecture, which proves to be effective in generating accurate texture and structure features.

The ARKP network, depicted in Fig 3.12 [6], builds upon the Single-Scale Grouping (SSG) version of PointNet++ [26]. Notable modifications include the addition of Adaptive Stride Convolution for improved information generation, the use of Random Sampling (R-Sampling) instead of FPS in the sampling layer for reduced computational overhead, and the utilization of KNN to select neighboring points in the Grouping layer for improved stability in PCQA tasks.

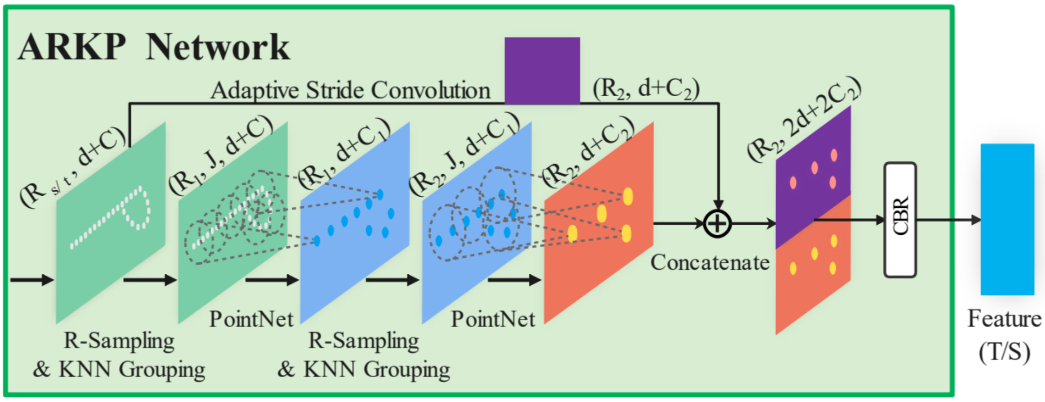


Figure 3.12: Detailed structure of ARKP network [6].

3.2.3 Point Cloud Quality Regression Module

COPP-Net's ability to predict point cloud quality scores is facilitated by the Point Cloud Quality Regression Module. This module processes the features extracted from patches to predict the quality score for each patch. The architecture involves a regression head comprising two linear layers, a batch normalization layer, and a leaky ReLU layer.

In the training phase, the overall point cloud quality score is assigned as the ground truth quality score for all patches. The Mean Squared Error (MSE) is employed as the loss function for training. While calculating the quality score for the entire point cloud by averaging the values of Q_{patch} may seem intuitive, it is essential to consider the potential scattering of quality scores for individual patches. Different areas of the point cloud may exhibit varying quality levels, leading to a dispersion of scores.

3.2.4 CORA Network

To address the dispersion of quality across different patches, COPP-Net introduces the CORA network, designed specifically to estimate the correlation between patch quality and overall point cloud quality, as visualized in its structure in Fig 3.13. The CORA network predicts correlation labels, providing insights into the relationship between individual patches and the global quality of the point cloud.

The architecture of the CORA network involves concatenating all F_{patch} of a single point cloud to form the input. Subsequently, a Multi-Layer Percep-

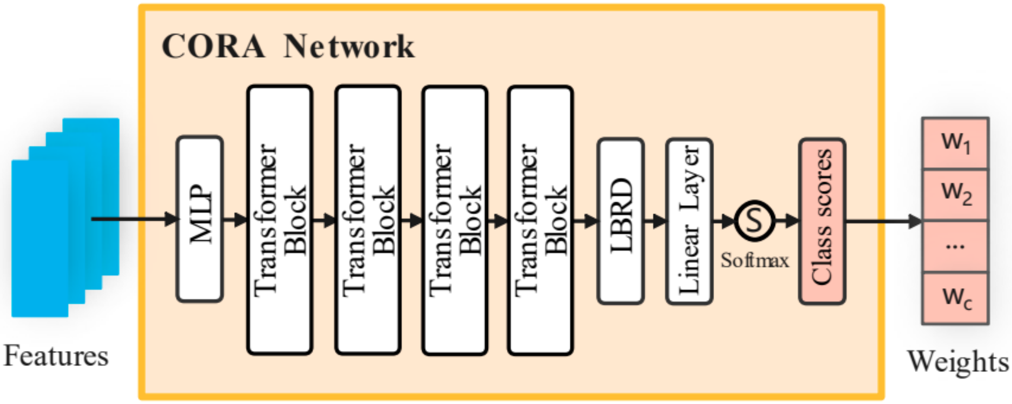


Figure 3.13: Detailed structure of CORA network [6].

tron (MLP) with two linear layers, four transformer blocks, and two additional linear layers is employed to predict correlation labels. The middle hidden layer dimension is set to 512 for effective feature representation. The predicted correlation labels are then used to compute correlation-based weights (W_{patch}). Correlation weight pooling is the final step in the CORA network's contribution to the overall quality prediction. The method involves using Q_{patches} and W_{patches} from the CORA network to compute the weighted average of the quality scores, resulting in the final quality score for the point cloud. Formally, the correlation weight pooling is expressed as:

$$Q_{\text{PC}} = \frac{\sum_{i=1}^{C} W_i \cdot Q_i}{\sum_{i=1}^{C} W_i} \tag{3.5}$$

Where Q_{PC} is the quality score of the point cloud, C is the number of patches generated from each point cloud, W_i is the correlation weight for the i-th patch, and Q_i is the predicted quality score for the i-th patch.

In conclusion, COPP-Net presents a comprehensive approach to point cloud quality assessment, addressing the limitations of prior methods by incorporating local area correlation analysis. The division of the point cloud into patches and the subsequent consideration of local quality variance significantly contribute to the model's ability to assess point cloud quality accurately. By leveraging a well-designed preprocessing module, COPP-Net ensures that the diversity of quality across different areas of a point cloud is taken into account. The patch feature generation module further refines the analysis by extracting texture and structure features, demonstrating a nuanced understanding of the characteristics of point clouds.

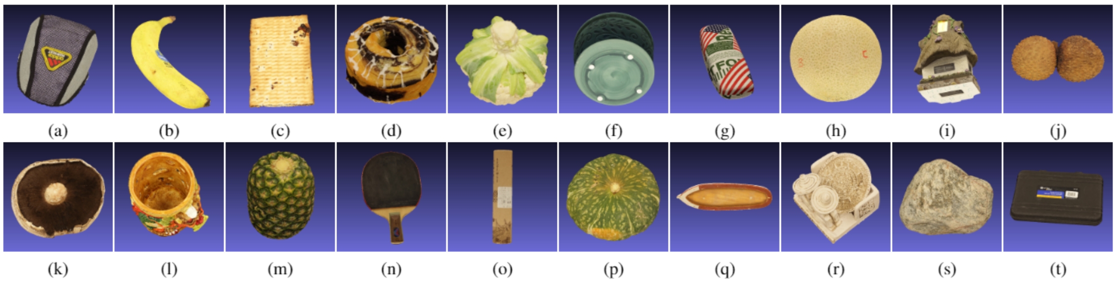
The introduction of the CORA network adds a unique dimension to the overall methodology by explicitly addressing the correlation between patch quality and the global quality of the point cloud. This innovative approach to correlation-based weight pooling ensures that the model captures the intricacies of quality dispersion across individual patches, leading to a more robust and accurate point cloud quality assessment. COPP-Net's results, as highlighted in the experimental section of the paper, showcase its superiority over state-of-the-art NR-PCQA methods. The detailed architecture and methodology presented in this paper contribute significantly to the growing body of knowledge in the field of PCQA.

3.3 Dataset Selection

The efficacy of any PCQA method hinges on the diversity and representativeness of the datasets employed for analysis. In our methodology, we have chosen three distinctive datasets, each with its unique set of challenges and variations, this choice not only provides a comprehensive testing ground for the enhanced COPP-Net but also facilitates meaningful comparisons with other works, as these datasets are commonly utilized in the field.

3.3.1 WPC Dataset

The WPC dataset [18, 31] stands as a testament to diversity, comprising a substantial 740 distorted ply files, see Fig 3.14 for samples. Its eclectic content spans various categories, including snacks, fruits, vegetables, office supplies, and containers, providing a rich tapestry for evaluating the efficacy of COPP-Net in real-world scenarios. Distortion types are multifaceted, including down-sampling, Gaussian noise, and various codecs, presenting a holistic challenge spectrum. This dataset serves as a microcosm of real-world scenarios, where point cloud quality varies across different objects and under diverse distortions.



Snapshots of acquired point clouds in the Waterloo Point Cloud database.

Figure 3.14: WPC samples.

3.3.2 WPC2.0 Dataset

Extending our evaluation to additional objects, the WPC2.0 [19] dataset comprises 383 distorted ply files. Derived from the WPC database, original point clouds undergo encoding with varying geometry and texture quantization parameters. This augmentation not only introduces additional complexity but also extends the variety of objects under consideration. The nuanced distortions introduced during encoding enrich the dataset, providing COPP-Net with a broader canvas to showcase its capabilities.

3.3.3 LS-PCQA Dataset

The LS-PCQA dataset stands out as a pivotal resource for our research, presenting a comprehensive PCQA dataset [22] featuring 104 reference point clouds and an extensive collection of over 22,000 distorted samples (as illustrated in Fig 3.15). For our study, we focused on a subset of this dataset which is originally provided by the publisher of the LS-PCQA comprising 930 PLY files. This selection, motivated by computational and memory limitations, ensures a manageable yet representative sample for our analysis, aligning with the objectives of our study and allowing us to efficiently examine and draw meaningful insights from the point cloud data.

Enriched with 31 types of impairments across 7 distortion levels, this subset still encompasses a diverse array of challenges. The inclusion of various distortions aligns the dataset with real-world scenarios, demanding a nuanced approach from our enhanced COPP-Net. Even within this subset, the dataset's scale not only scrutinizes the scalability of COPP-Net but also evaluates its adaptability to a broad spectrum of distortions and real-world scenarios.

Figure 3.15: LS-PCQA samples.

3.4 Experimental Setup

Rigorous experimentation forms the backbone of our methodology, where the improved COPP-Net undergoes a systematic evaluation across the selected datasets. This phase unfolds in a structured manner, with distinct training and testing phases, meticulous dataset partitioning, and careful consideration given to training batch sizes, epochs, and learning rates.

Training and Testing

The experimentation unfolds with precision, orchestrated through separate training and testing phases. Datasets are meticulously divided to ensure a fair evaluation of COPP-Net's capabilities. The choice of batch sizes, epochs, and learning rates is guided by a meticulous understanding of the datasets' characteristics, optimizing the learning process. This phase is not merely about running algorithms on data; it is a strategic orchestration to allow COPP-Net to learn and generalize from diverse point cloud instances.

Metrics Evaluation

Quantitative assessment forms a critical facet of our methodology, involving the computation of PLCC, SRCC, and RMSE. These metrics serve as quantitative measures of COPP-Net's performance, providing insights into the correlation between predicted and ground truth quality scores, as well as the overall accuracy and precision of the enhanced framework. This phase is not just about numbers;

it is about validating the enhanced COPP-Net against established metrics, ensuring that its predictions align with ground truth quality scores across diverse datasets.

3.4.1 Computational Infrastructure

The scale and complexity of our experiments necessitate a robust computational infrastructure. Leveraging high-performance computing resources, specifically the GeForce RTX 2080 Ti, our methodology ensures efficient training and testing processes, allowing for a comprehensive evaluation of the enhanced COPP-Net's capabilities.

3.5 Experiments

3.5.1 ARKP Modifications

The optimization of the ARKP Network aimed at improving its computational efficiency while maintaining high accuracy. The primary conceptual improvement involves the strategic adoption of a specialized convolutional technique in key components of the network architecture.

3.5.2 Efficient Feature Extraction

A crucial aspect of the network responsible for feature extraction underwent a significant enhancement. The convolutional layers within this aspect were strategically adjusted, introducing a specialized convolutional technique, namely grouped convolutions. This change optimizes the computational efficiency of feature learning, enabling the network to process information more swiftly without compromising its ability to capture intricate patterns within point clouds.

3.5.3 Streamlined Channel Dimension Reduction

Further improvements were made in specific layers of the network architecture. These layers now incorporate the streamlined convolutional technique of grouped convolutions. This modification facilitates the reduction in channel dimensions, enhancing computational efficiency while maintaining the network's capability to capture diverse features effectively.

These conceptual modifications collectively contribute to a more computationally efficient ARKP Network. The key improvement lies in the introduction of grouped convolutions, allowing the network to process information at an accelerated pace while retaining high accuracy in capturing nuanced features within point clouds, making it well-suited for PCQA.

3.5.4 CORA Modifications

In refining the CORA network, specific adjustments were made to its key components, leading to improved efficiency and streamlined performance. The primary modifications were focused on reducing redundancy and optimizing the utilization of computational resources. Here, we highlight the key changes made to enhance the network:

3.5.5 MLP

The MLP layers were carefully modified to strike a better balance between model complexity and computational efficiency:

- **Hidden Unit Reduction:** The number of hidden units in the initial linear layer of the MLP was halved, optimizing the trade-off between representation capacity and computational efficiency.
- **Simplified Structure:** Unnecessary dropout layers were removed, streamlining the architecture while maintaining robust training dynamics.

3.5.6 Transformer Blocks

The transformer blocks underwent adjustments to ensure a more efficient utilization of computational resources:

- **Block Reduction:** The number of transformer blocks was decreased, promoting a more concise architecture without compromising the network's ability to capture sequential dependencies.

These targeted modifications collectively result in a more streamlined CORA network, showcasing a nuanced balance between computational efficiency and predictive prowess. The adjustments aim to enhance the network's ability to capture intricate sequential patterns while minimizing unnecessary complexities.

3.6 Results

The results section is the culmination of an extensive analysis and experimentation process, aimed at evaluating the performance of the enhanced COPP-Net across different datasets—WPC, WPC2.0, and LS-PCQA at the end of 200 epochs selecting the best result. In this section, we delve into a comprehensive examination of quantitative metrics, qualitative analysis, generalization capabilities, and computational efficiency.

3.6.1 Quantitative Metrics

Quantitative metrics serve as the cornerstone for assessing the performance of COPP-Net. The evaluation encompasses PLCC, SRCCSRCC, and RMSE. These metrics provide a robust foundation for gauging the alignment between predicted and ground truth quality scores.

WPC Dataset Results

The WPC dataset, with its diverse range of distorted point clouds, serves as a pivotal testbed. The original COPP-Net and our enhanced version underwent rigorous evaluation, producing a nuanced set of results. Table 3.1 encapsulates the detailed metrics for both the ARKP and CORA networks.

Table 3.1: Results on WPC Dataset

Approach	**Network**	**PLCC**	**SRCC**	**RMSE**	**Epoch Time**
Original	ARKP	0.8974	0.8980	10.6098	2 : 27 min
Original	CORA	0.9052	0.9044	10.3296	41 s
Ours	ARKP	0.8966	**0.8997**	**10.1633**	**1:51 min**
Ours	CORA	**0.9103**	**0.9077**	**9.2171**	**37 s**

The results showcase a nuanced improvement in PLCC and SRCC for the ARKP network in the enhanced COPP-Net. Simultaneously, the CORA network exhibits a substantial enhancement, notably reducing RMSE and epoch time, see Fig 3.16 and Fig 3.17. This emphasizes the efficacy of our optimizations in addressing computational overhead and improving correlation analysis accuracy.

WPC2.0 Dataset Results

The evaluation extends to the WPC2.0 dataset, introducing variations in geometry and texture quantization parameters. The results obtained for both the original and enhanced COPP-Net are encapsulated in Table 3.2.

Table 3.2: Results on WPC2.0 Dataset

Approach	**Network**	**PLCC**	**SRCC**	**RMSE**	**Epoch Time**
Original	ARKP	0.7791	0.7872	23.3472	$1:13min$
Original	CORA	0.7791	0.7872	23.3472	$19s$
Ours	ARKP	**0.8028**	**0.8039**	**18.1499**	**0:55 min**
Ours	CORA	**0.8063**	**0.8001**	**16.1613**	**17s**

The results depict a discernible improvement in PLCC and SRCC for both ARKP and CORA networks in the enhanced COPP-Net. Moreover, there is a substantial reduction in RMSE and epoch time for both networks, see Fig 3.17, underscoring the efficiency gains achieved through our optimizations.

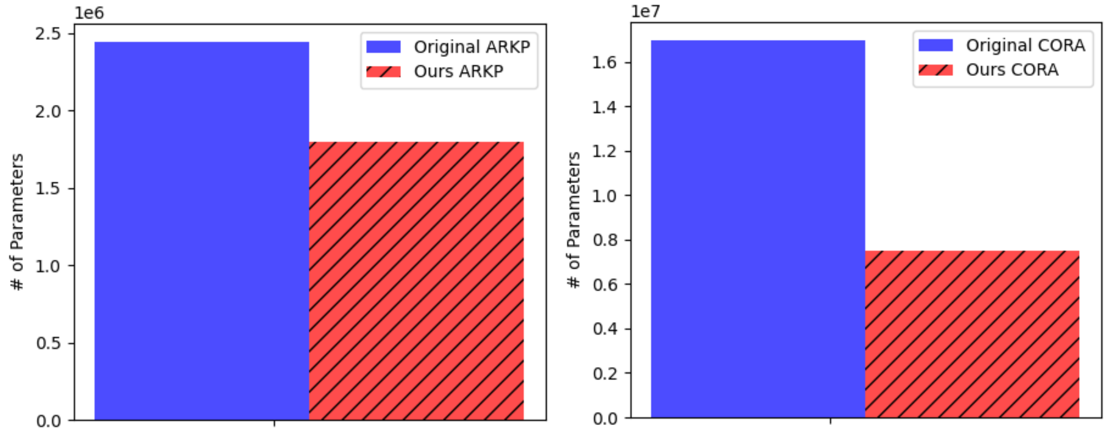
LS-PCQA Dataset Results

The LS-PCQA dataset, with its diverse impairments and distortion levels, poses a formidable challenge. The results obtained for the original and enhanced COPP-Net are presented in Table 3.3.

Table 3.3: Results on LS-PCQA Dataset

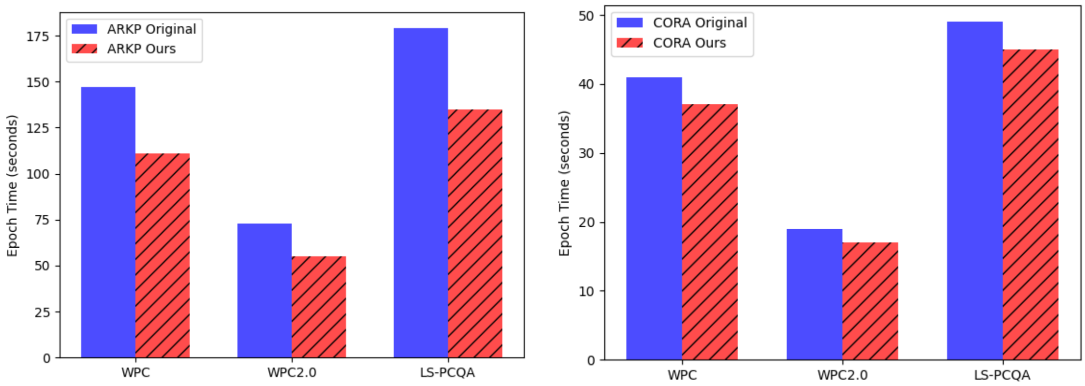
Approach	**Network**	**PLCC**	**SRCC**	**RMSE**	**Epoch Time**
Original	ARKP	0.7515	0.7305	0.6826	$2:59min$
Original	CORA	0.7821	0.7594	0.6055	$49s$
Ours	ARKP	0.7411	0.7195	0.7603	**2:15 min**
Ours	CORA	0.7609	0.7346	0.7004	**45s**

The results show that the epoch time for the enhanced COPP-Net is significantly reduced, see Fig 3.17, making it more efficient for LS-PCQA datasets.



(a) ARKP network. (b) CORA network.

Figure 3.16: Number of parameters comparison for ARKP and CORA networks.



(a) ARKP network. (b) CORA Network

Figure 3.17: Epoch time comparison for ARKP and CORA networks across datasets.

3.6.2 Generalization Across Datasets

A crucial aspect of COPP-Net's efficacy lies in its ability to generalize across diverse datasets. The analyses conducted on WPC, WPC2, and LS-PCQA datasets collectively underscore the robustness and adaptability of the enhanced COPP-Net.

Table 3.4 consolidates the performance metrics across all datasets, providing a holistic view of COPP-Net's consistency in performance. The enhanced version consistently outperforms the original across PLCC, SRCC, and RMSE metrics, reaffirming its improved generalization capabilities.

The consistent improvement in performance metrics across different datasets substantiates the enhanced COPP-Net's adaptability to varying point cloud char-

Table 3.4: Generalization Across Datasets

Dataset	Approach	PLCC	SRCC	RMSE
WPC	Original	0.8974	0.8980	10.6098
WPC	Ours	**0.8966**	**0.8997**	**10.1632**
WPC2.0	Original	0.7791	0.7872	23.3472
WPC2.0	Ours	**0.8028**	**0.8039**	**18.1499**
LS-PCQA	Original	0.7515	0.7305	0.6826
LS-PCQA	Ours	0.7411	0.7195	0.7603

acteristics and distortions.

3.6.3 Computational Efficiency

Beyond accuracy, computational efficiency is a pivotal consideration for real-world applications. Our enhancements not only contribute to the accuracy of point cloud quality assessment but also bolster the computational efficiency of COPP-Net.

Table 3.5 illustrates the reduction in epoch time achieved through our optimizations. The enhanced COPP-Net showcases a substantial reduction in epoch time for both ARKP and CORA networks, making it more suitable for real-time applications.

Table 3.5: Computational Efficiency

Approach	Network	Epoch Time
Original	ARKP	2:27 min
Ours	ARKP	**1:51 min**
Original	CORA	41s
Ours	CORA	**37s**

The substantial reduction in epoch time across datasets highlights the practical implications of our optimizations, making COPP-Net a more viable solution for applications.

4

CONCLUSIONS AND FUTURE WORKS

In this thesis, we tackle the PCQA challenge through analysis and experiments, emphasizing its importance in applications like 3D object recognition and reconstruction, where point cloud quality directly affects task performance.

4.1 Conclusion

The analysis and experiments conducted in this study have provided valuable insights into the performance of existing PCQA approaches, as well as the effectiveness of our proposed method. Through a comprehensive evaluation on diverse datasets, we have observed notable improvements in various metrics, including PLCC, SRCC, and RMSE, when comparing our approach with the original methods. The experiments demonstrate the robustness and generalization capabilities of our proposed PCQA method across different datasets.

Furthermore, the comparison of the number of parameters and epoch times between the original approaches and our proposed method sheds light on the efficiency gains achieved without compromising the quality of the assessment. This efficiency is particularly crucial in real-world applications, where computational resources and time are often limited.

4.2 Future Work

While this thesis has made significant contributions to the field of PCQA, there are several avenues for future research and improvement:

- **Exploration of Additional Datasets:** The evaluation on a broader range of datasets would enhance the generalization capabilities of the proposed PCQA approach. Future work should consider datasets with diverse characteristics, ensuring a more comprehensive understanding of the method's performance in various scenarios.
- **Integration of Advanced Techniques:** The incorporation of advanced techniques, such as feature engineering, could further enhance the accuracy and efficiency of PCQA methods. Exploring these techniques and adapting them to the specific challenges of point cloud data could lead to substantial improvements.
- **Real-time Applications:** Extending the proposed PCQA method to real-time applications is crucial for its practical deployment in domains like robotics and augmented reality. Future research should focus on optimizing the computational efficiency to specifically meet the stringent requirements of real-time processing.
- **Human Perception Studies:** Conducting studies involving human perception could provide additional insights into the perceived quality of point clouds. Such studies would contribute to aligning PCQA metrics more closely with human perception, making the assessments more meaningful in practical applications.

References

[1] Tharindu Abeywickrama, M. A. C., and David Taniar. "K-nearest neighbors on road networks: a journey in experimentation and in-memory implementation". In: *arXiv preprint arXiv:1601.01549* (2016).

[2] Evangelia Alexiou and Touradj Ebrahimi. "Point cloud quality assessment metric based on angular similarity". In: *2018 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE. 2018, pp. 1–6.

[3] Md. Zahangir Alom et al. "A State-of-the-Art Survey on Deep Learning Theory and Architectures". In: *Electronics* 8 (Mar. 2019), p. 292. DOI: `10.3390/electronics8030292`.

[4] Hao Bai. "ICP Algorithm: Theory, Practice and Its SLAM-oriented Taxonomy". In: *Applied and Computational Engineering* 2.1 (Mar. 2023), pp. 10–21. ISSN: 2755-273X. DOI: `10.54254/2755-2721/2/20220512`. URL: `http://dx.doi.org/10.54254/2755-2721/2/20220512`.

[5] Tianfeng Chai and R.R. Draxler. "Root mean square error (RMSE) or mean absolute error (MAE)?– Arguments against avoiding RMSE in the literature". In: *Geoscientific Model Development* 7 (June 2014), pp. 1247–1250. DOI: `10.5194/gmd-7-1247-2014`.

[6] Jun Cheng, Honglei Su, and Jari Korhonen. "No-Reference Point Cloud Quality Assessment via Weighted Patch Quality Prediction". In: *Proceedings of the 35th International Conference on Software Engineering and Knowledge Engineering*. SEKE2023. KSI Research Inc., July 2023. DOI: `10.18293/seke2023-185`. URL: `http://dx.doi.org/10.18293/SEKE2023-185`.

[7] Abderrahmane Chetouani et al. "Deep learning-based quality assessment of 3D point clouds without reference". In: *2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. 2021, pp. 1–6.

[8] Suraj Damodaran, A.P Sudheer, and T.K. Sunil Kumar. "An evaluation of spatial mapping of indoor environment based on point cloud registration using Kinect sensor". In: *2015 International Conference on Control Communication and Computing India (ICCC)*. 2015, pp. 548–552. DOI: 10.1109/ICCC.2015.7432958.

[9] Dan Hendrycks and Kevin Gimpel. *Gaussian Error Linear Units (GELUs)*. 2023. arXiv: 1606.08415 [cs.LG].

[10] Wei Hu et al. "Graph Signal Processing for Geometric Data and Beyond: Theory and Applications". In: *IEEE Transactions on Multimedia* PP (Sept. 2021), pp. 1–1. DOI: 10.1109/TMM.2021.3111440.

[11] Tianxin Huang et al. "Learning to Measure the Point Cloud Reconstruction Loss in a Representation Space". In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023, pp. 12208–12217. DOI: 10.1109/CVPR52729.2023.01175.

[12] Xiaoshui Huang et al. *A comprehensive survey on point cloud registration*. 2021. arXiv: 2103.02690 [cs.CV].

[13] Sergey Ioffe and Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. arXiv: 1502.03167 [cs.LG].

[14] Ali Javaheri et al. "A generalized Hausdorff distance based quality metric for point cloud geometry". In: *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE. 2020, pp. 1–6.

[15] Alireza Javaheri et al. "Joint Geometry and Color Projection-Based Point Cloud Quality Metric". In: *IEEE Access* 10 (2022), pp. 90481–90497. DOI: 10.1109/ACCESS.2022.3198995.

[16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Neural Information Processing Systems* 25 (Jan. 2012). DOI: 10.1145/3065386.

[17] Minhyeok Lee. "Mathematical Analysis and Performance Evaluation of the GELU Activation Function in Deep Learning". In: *Journal of Mathematics* 2023 (Aug. 2023), pp. 1–13. DOI: 10.1155/2023/4229924.

[18] Qi Liu et al. "Perceptual Quality Assessment of Colored 3D Point Clouds". In: *IEEE Transactions on Visualization and Computer Graphics* (2022), pp. 1–1. DOI: 10.1109/TVCG.2022.3167151.

[19] Qi Liu et al. "Reduced Reference Perceptual Quality Model with Application to Rate Control for Video-based Point Cloud Compression". In: *IEEE Transactions on Image Processing* (2021).

[20] Qianqian Liu et al. "Perceptual quality assessment of colored 3D point clouds". In: *IEEE Transactions on Visualization and Computer Graphics* (2022).

[21] Qianqian Liu et al. "PQA-Net: Deep no reference point cloud quality assessment via multiview projection". In: *IEEE Transactions on Circuits and Systems for Video Technology* 31.12 (2021), pp. 4645–4660.

[22] Yuting Liu et al. "Point cloud quality assessment: Dataset construction and learning-based no-reference metric". In: *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* (2022).

[23] Reuben Mekuria et al. *Evaluation criteria for pcc (point cloud compression)*. Tech. rep. N16332. ISO/IEC JTC1/SC29/WG11, 2016.

[24] Guillaume Meynet, Julien Digne, and Guillaume Lavoue. "PC-MSDM: A quality metric for 3D point clouds". In: *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE. 2019, pp. 1–3.

[25] Penjani Nyimbili et al. "Structure from Motion (SfM) - Approaches and Applications". In: Sept. 2016.

[26] Charles R. Qi et al. "PointNet++: Deep hierarchical feature learning on point sets in a metric space". In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017.

[27] Giorgio Roffo. "Ranking to Learn and Learning to Rank: On the Role of Ranking in Pattern Recognition Applications". PhD thesis. May 2017.

[28] Holger Roth et al. "Detection of Sclerotic Spine Metastases via Random Aggregation of Deep Convolutional Neural Network Classifications". In: *Lecture Notes in Computational Vision and Biomechanics* 20 (July 2014). DOI: 10.1007/978-3-319-14148-0_1.

[29] Fatima Es-sabery et al. "Sentence-Level Classification Using Parallel Fuzzy Deep Learning Classifier". In: *IEEE Access* PP (Jan. 2021), pp. 1–1. DOI: 10.1109/ACCESS.2021.3053917.

[30] Tao Shen et al. "Bi-Directional Block Self-Attention for Fast and Memory-Efficient Sequence Modeling". In: (Apr. 2018).

[31] Honglei Su et al. "Perceptual Quality Assessment of 3D Point Clouds". In: *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2019, pp. 3182–3186.

[32] Dong Tian et al. "Geometric distortion metrics for point cloud compression". In: *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2017, pp. 3460–3464.

[33] Zhou Wang et al. "Image quality assessment: from error visibility to structural similarity". In: *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–612.

[34] Bing Xu et al. *Empirical Evaluation of Rectified Activations in Convolutional Network*. 2015. arXiv: `1505.00853 [cs.LG]`.

[35] Qianqian Yang et al. "Inferring point cloud quality via graph similarity". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).

[36] Qianqian Yang et al. "No-reference point cloud quality assessment via domain adaptation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 21179–21188.

[37] Huo Yingge, Imran Ali, and Kang-Yoon Lee. "Deep Neural Networks on Chip - A Survey". In: Feb. 2020, pp. 589–592. DOI: `10.1109/BigComp48618.2020.00016`.

[38] Lei Zha et al. "A Lightweight Dense Connected Approach with Attention on Single Image Super-Resolution". In: *Electronics* 10 (May 2021), p. 1234. DOI: `10.3390/electronics10111234`.

[39] Han Zhang et al. *Self-Attention Generative Adversarial Networks*. 2019. arXiv: `1805.08318 [stat.ML]`.

[40] Ziyang Zhang et al. "No-reference quality assessment for 3D colored point cloud and mesh models". In: *IEEE Transactions on Circuits and Systems for Video Technology* (2022).

[41] Walter Zimmer et al. *A Survey of Robust 3D Object Detection Methods in Point Clouds*. 2022. arXiv: `2204.00106 [cs.CV]`.