Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 13, No. 1, 2024

A new method of image recognition based on deep learning generates adversarial networks and integrates traditional algorithms

Yutian Yang[1, a], Zexi Chen[2, b], Yafeng Yan[3, c], Muqing Li[4, d], Tana Gegen[5, e]

[1] University of California, Davis, USA
[2] North Carolina State University, USA
[3] Stevens Institute of Technology, USA
[4] University of California San Diego, USA
[5] Johns Hopkins University, USA

[a]yytyang@ucdavis.edu, [b]chenzx.hit09@gmail.com, [c]yanyafeng0105@gmail.com, [d]MUL003@ucsd.edu, [e]tanagegen7@gmail.com

Abstract: This paper discusses an innovative image recognition method, which combines the advanced feature learning ability of generative adversarial networks (Gan) with the robustness of traditional image recognition algorithms. Based on small-sample training, Gan can generate high-quality image samples to expand the training set. Therefore, by designing and training a generative adversarial network, the real image data and its labels can be integrated into the training set. The GAN-generated images, together with corresponding labels, are input into SVM for training, and appropriate kernel functions and parameters are selected to optimize the SVM model to maximize classification performance. GAN is good at data generation and feature learning, while SVM is good at problems with clear classification boundaries, and this model combining the two methods is used in image recognition.

Keywords: SVM; GAN; Kernel function.

1. Introduction

1.1. Introduce

With the swift advancement of artificial intelligence technology, image recognition as one of its core applications, generative adversarial network (GAN) has attracted wide attention for its excellent performance in generating realistic images and learning deep features. At the same time, the esteemed machine learning methodology, The Support Vector Machine (SVM) persists as a potent means for addressing classification tasks, attributed to its notable efficacy and strength in partitioning high-dimensional data with steadfastness. This paper aims to explore a novel image recognition framework that cleverly combines the innovative feature learning potential of GAN with the robust classification capability of SVM to achieve a more efficient image recognition system.

A Generative Adversarial Network (GAN) represents a sophisticated deep learning architecture, comprised of dual components: the Creator and the Identifier. The Creator's objective revolves around synthesizing plausible image instances from arbitrary noise, while the Identifier's mission is to segregate these fabricated samples from actual images. Through this cat-and-mouse game, the two networks compete and co-evolve, eventually allowing the generator to output images that are almost indistinguishable from real data. More importantly, in the process, the generator learns to capture and express the intrinsic feature structure of the dataset, which is extremely valuable for subsequent image recognition tasks. Support vector machines (SVMS), as a supervised learning model, achieve classification by finding the optimal hyperplane in a data set, and are especially suitable for processing linearly separable or approximately linearly separable data. The key of SVM is its maximum interval principle, which ensures that the model can not only accurately classify training data, but also has good generalization ability. By introducing kernel techniques, SVM can even effectively deal with nonlinear problems, further expanding its application range. In the image recognition scene, SVM can accurately classify different types of images according to the extracted features, especially when the boundary between classes is clear, SVM performance is particularly outstanding.

In view of the strong ability of GAN in feature learning and the robustness of SVM in classification tasks, we designed and trained a GAN model using a large amount of real image data and its corresponding labels as the training basis. Through this process, the generator not only learns how to generate new images, but also implicitly learns the distribution of image features. We then enter the real image data and its labels, together with the synthetic image generated by the generator and the corresponding labels, into the SVM for training. This method not only enriches the diversity of training set, but also provides more potential classification boundary information for SVM, which helps the model to find more accurate decision boundary in complex feature space. In order to take full advantage of this framework, choosing the right kernel function and tuning parameters for an SVM model is crucial. The choice of kernel function directly affects the ability of the model to deal with nonlinear problems, while parameter adjustment determines the generalization ability and noise sensitivity of the model. By evaluating and adjusting the model, we can systematically explore these parameter Spaces and find the optimal configuration to maximize classification performance.

1.2. Related work

The novel approach of integrating generative adversarial networks (GANs) with Support Vector Machines (SVMs) for image recognition builds upon a substantial body of research that explores the synergy between deep learning's innovative capabilities and the robustness of traditional algorithms. Xu et al.[1] and Zhang et al. [2] have made significant contributions to the field by enhancing image recognition with multimodal deep learning and multi-scale convolutional neural networks, respectively, highlighting advancements that align with our method's goals of optimizing image processing and enhancing training datasets with GAN-generated images. Similarly, Wang et al. [3] and Xiao et al.[4] demonstrate practical applications of these technologies in medical image processing, from automated report generation to specific cancer cell classification, which are particularly relevant given our focus on enhancing image recognition accuracy and reliability in complex scenarios.

Additionally, the work of Dai et al. [5] addresses the critical issue of unintended biases in AI systems using LSTM and attention-based models, a concern that directly relates to our use of GANs and SVMs where ensuring unbiased synthetic image generation is crucial. Efficiency optimizations in deep learning processes are explored by Mei et al. [6], while Lu et al.[7] delve into the scalability and privacy preservation of deep learning applications through federated learning, both of which provide insights into potential enhancements for our framework.

Furthermore, the importance of precise feature extraction and selection is emphasized by Liu and Song [8], and the detailed image segmentation techniques employed by Zhao et al. [9] showcase the application of machine learning in achieving high precision in image analysis tasks. These studies bolster the methodological underpinnings of our approach, suggesting that effective feature handling is paramount for robust classification outcomes.

In exploring the broader applications and implications of deep learning in medical diagnostics, Yan et al. [10] and Wang et al. [11] provide examples of how neural networks are being used to predict patient outcomes and diagnose complex conditions such as Alzheimer's disease, further illustrating the expansive potential of combining GANs with SVMs. These references collectively provide a rich tapestry of research that supports and contextualizes our proposed method, highlighting its potential to push the boundaries of image recognition technology by leveraging both generative and discriminative machine learning models. This integrated approach promises not only to enhance the accuracy and efficiency of image recognition systems but also to expand their applicability across various challenging domains.

To sum up, image super resolution reconstruction technology has practical application value in various fields. Researchers usually improve the image resolution from both hardware and software, but the use of hardware equipment often brings high costs and high technical risks, while improving the image resolution through algorithms can effectively reduce costs and avoid technical risks. Therefore, image super resolution reconstruction technology has greater research significance and wide application prospects in the development and application.

2. Theoretical model

2.1. SVM model

Support vector machine is a binary classification algorithm based on integrated small sample statistical analysis. It skillfully uses kernel function techniques to overcome the "dimensional disaster" problem and nonlinear classification problems, and can maintain computational efficiency even in high-dimensional space mapping and avoid complexity inflation. Therefore, it shows a high degree of application flexibility and excellent classification accuracy in many fields. Its core mechanism revolves around the construction of linear classification bounds with maximum structured interval, and through the implantation of kernel method, it is cleverly transformed into nonlinear classifier, which greatly enhances the generalization performance of the model to new data. The core focus of SVM is to minimize the empirical risk and uncertainty interval, while pursuing the maximum expansion of the interval to ensure that stable statistical regularity can be captured even in the face of limited samples. Specifically, the learning path of SVM is designed to identify hyperplanes that can perfectly distinguish amidst training datasets, while optimizing the geometric separation among data points. As illustrated in Figure 1, the central solid line $w \cdot x + b = 0$ represents such a classification hyperplane, successfully splitting the dataset into two parts. In the case of completely linearly divisible data, although there exists an uncountable multitude of these dividing hyperplanes, only The hyperplane boasting the greatest geometric margin is distinct.

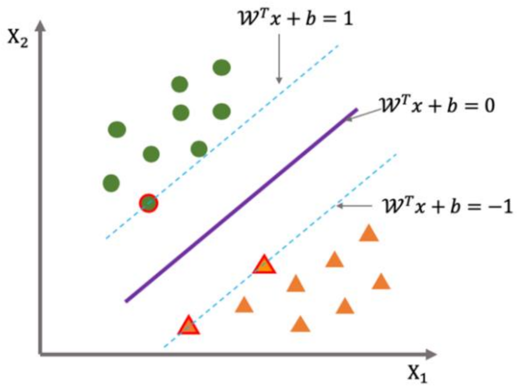

Figure 1. SVM classifier

In practice, there are many data that cannot be divided by linear functions in two-dimensional space, but can be easily divided by hyperplanes in higher dimensional space. Generally, problems of nonlinear classification within a low-dimensional realm are converted into linear classification tasks within an augmented high-dimensional feature space via the application of a nonlinear projection transformation mechanism. Enabling classification through learning a linear support vector machine within the fabricated feature space, this approach harnesses the dual aspect of linear SVM learning. Here, both the pursuit of the objective and the formulation of the classification rule hinge solely on the internal product of examples. Consequently, articulating the nonlinear transformation isn't mandatory; the process circumvents this by substituting the inner product with a kernel function, effectively realizing an implicit elevation of data to more complex dimensions.

The kernel function is characterized through the inner product of a pair of instances following a nonlinear transformation. If K(x,z) is a function, and should a transformation from the original input space exist, to the feature space $\phi(x)$, then for any input space x,z, there is:

$$K(x,z) = \phi(x) \cdot \phi(z) \quad (1)$$

Then the kernel function is substituted, and the representation of a nonlinear support vector machine is fashioned as:

$$f(\mathrm{x}) = \mathrm{sign}\left(\sum_{\mathrm{i}=1}^{N} \alpha_{\mathrm{i}}^{*} y_i K(x, x_i) + b^*\right) \quad (2)$$

The procedure for the support vector machine learning methodology designed for nonlinear scenarios unfolds as follows:

Input: The training data set is : $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, in its $\mathrm{x_i} \in \mathbb{R}^n$, $\mathbf{y_i} \in \{+\mathbf{1}, -\mathbf{1}\}, \mathbf{i} = \mathbf{1}, \mathbf{2}, \dots \mathbf{N}$.

Output: Hyperplane segregation and classification decision function.

Identifying an apt kernel function, denoted as K(x, z), alongside a penalty coefficient, C, greater than zero, facilitates the formulation and resolution of a convex quadratic programming challenge:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \mathrm{y}_i \mathrm{y}_j \mathrm{K}(\mathrm{x}_i, \mathrm{x}_j) - \sum_{i=1}^{N} \alpha_i \mathrm{s.}t. \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$0 \le \alpha_i \le C, \mathrm{i} = 1,2, \dots, N \quad (3)$$

Solve equation (3) to obtain the optimal solution:

$$\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T \quad (4)$$

Substitute into formula (3) and choose an element from $\alpha *$, $\alpha *$ to satisfy the condition $0 < \alpha * < \mathrm{C}$

Calculation:

$$\boldsymbol{b}^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i K(x_i, x_j) \quad (5)$$

To find the classification decision function:

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_i^* y_i K(x, x_i) + b^*\right) \quad (6)$$

The commonly used kernel function, Gaussian kernel function, is defined as:

$$K(x, z) = exp\left(-\frac{\| x - z \|^2}{2\sigma^2}\right) \quad (7)$$

For this scenario, the function determining classification is:

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_i^* y_i \, exp\left(-\frac{\| x - z \|^2}{2\sigma^2}\right) + b^*\right) \quad (8)$$

2.2. GAN network

Generating adversarial networks means creating artificial instances from a dataset that retain the characteristics of the original dataset to the maximum extent possible. A number of generative modeling frameworks based on these adversarial training principles, collectively known as generative adversarial networks, have been developed to address the challenge of expanding datasets and generating realistic data. Their uniqueness comes from their architecture (Figure 2), which incorporates two competing components: generators and discriminators. The generator aims to create new samples that approximate the actual data based on a preset probabilistic model, while the discriminator's job is to identify the differences between these generated instances and the members of the real data set and assess whether they conform to real-world data patterns. In the iterative training process, the two promote each other and continuously optimize: the generator strives to create more realistic and difficult to discern output to confuse the discriminator; Instead, the discriminator is learning more and more, sharpening the sharpness of distinguishing real data from synthetic data. This back-and-forth process constitutes a dynamic "hunt and escape" game, the ultimate goal of which is to force the generator to produce an output that closely resembles the original data and is barely realistic.

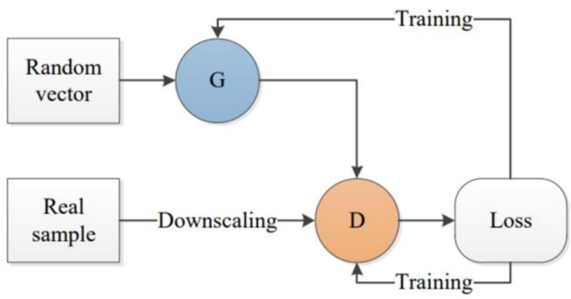

Figure 2. GAN architecture

The generator component, G, endeavors to produce increasingly authentic samples with the intent of deceiving the discriminator, D, into mistaking them as genuine inputs. Discriminator D, on the other hand, is constantly learning how to better detect the differences between generated and real samples, thereby improving its ability to distinguish. The generator (G) creates a simulated sample from a random vector, while the discriminator (D) receives both the generated sample and the real sample, attempting to differentiate between the two. In the course of training, both the generator G and the discriminator D progressively enhance their capabilities via recursive optimization cycles. The terminal aspiration is for the generator G to fabricate instances virtually identical to genuine data, thus rendering them barely distinguishable. This adversarial learning process helps Generator G to generate higher-quality simulation data.

2.3. GAN-SVM model

Gan-SVM model is a combination of GAN and SVM, aiming to integrate the powerful ability of GAN in data generation and the efficient performance of SVM in classification tasks. In traditional image recognition tasks, SVM is often used as a classifier, because it can effectively deal with high-dimensional feature space and find the best classification boundary. However, SVM performance is highly dependent on the quality of the extracted features. As a generative model, Gan can learn and simulate the distribution of training data to generate new and diversified samples. Combined with the two, Gan can be used to generate additional training data for boosting characteristic extraction and refining image discernment proficiency, as shown in Figure 3.

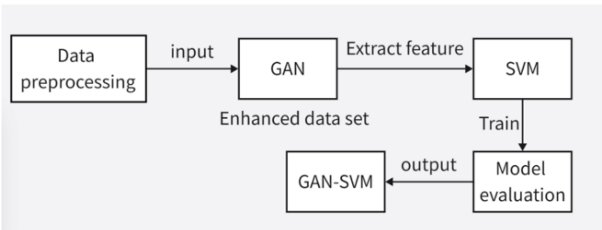

Figure 3. Gan-SVM model frame diagram

GAN-SVM first performs pre-processing operations such as cleaning and normalization of the original image data to prepare for the subsequent steps. Employing random noise vectors as its stimuli, the generator's objective centers on producing replicas that closely emulate true images. Concurrently, the discriminator undergoes training to distinguish bona fide images from those artifices crafted by the generator. In the course of instruction, the generator and discriminator engage in a competitive dynamic, resulting in the generator gradually improving its ability to generate increasingly realistic images. The GAN-generated images are used as additional training data to enhance the SVM training set.

3. Experimental analysis

Gan can generate a large number of diverse images, which can be used as data enhancement to help alleviate overfitting problems, especially in the case of scarce annotation data and the features generated by Gan are often more abstract and discriminative, that can enhance the efficacy of categorization models, especially when dealing with complex or high-dimensional image recognition tasks. By generating samples under different conditions, Gan can help SVM learn a wider range of patterns, thereby improving the model's performance on unseen data. The dataset was obtained from an open platform and was processed for training both GAN-SVM models and GAN models. The dataset utilizes Linked Data [12]techniques to amalgamate diverse data formats, thereby improving data interoperability and facilitating comprehensive analysis.These models were evaluated using accuracy, recall, and AUC metrics. The calculation methods for these metrics are shown in Equations (9-12).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (9)$$

$$recall = \frac{TP}{TP + FN}. \quad (10)$$

$$TPR = \frac{TP}{TP + FN} = \frac{TP}{P} \quad (11)$$

$$FPR = \frac{FP}{FP + TN} = \frac{FP}{N} \quad (12)$$

In all samples, Ac represents the accuracy of the model's predictions, Rec denotes the recall ratio, and AUC is a comprehensive measure of the model's classification ability. Accuracy (Ac) signifies the proportion of the model that is correctly classified among all the samples, and is a basic indicator to measure the comprehensive effectiveness of the model[13]. Recall (Rec) denotes the proportion of actual positive instances accurately recognized by the model among all samples, reflecting the sensitivity of the classifier towards authentic instances. AUC (Area Under Curve) refers to the region beneath the Receiver Operating Characteristic (ROC) curve, providing a more comprehensive classification performance evaluation by comparing the performance of models under different thresholds. Combining these three metrics allows a more comprehensive assessment of the model's strengths and weaknesses, as shown in Table 1.

Table 1. Evaluation of GAN-SVM model and GAN model under three indexes

Model	Ac	Rec	AUC
GAN	**87.64%**	**78.39%**	**0.92**
GAN-SVM	**93.42%**	**85.64%**	**0.96**

From the information shown in the table data, the accuracy of GAN model is 87.64%, the recall rate is 78.39%, and the area under the curve (AUC) score is 0.92. In contrast, the model combining GAN and SVM technology has improved in all indexes, with the accuracy rising to 93.42%, the recall rate reaching 85.64%, and the AUC score increasing to 0.96. This series of data shows that the GAN-SVM model has advantages over the basic GAN model in the three key evaluation dimensions of accuracy, recall rate and AUC. Especially noteworthy is the value of AUC, as a comprehensive index to measure the classification efficiency of the model, the increase of its value means the enhancement of classification performance. From this criterion, it is obvious that the combined model of GAN and SVM shows better performance characteristics.

Then, for further intuitive evaluation, ablation experiments were performed on GAN-SVM model using accuracy evaluation criteria, and separate GAN model and SVM model were used to train real data. Gan- SVM model represents the model proposed by us. SVM wo GAN Images represent the accuracy of SVM analysis after removing the images generated by GAN. GAN wo SVM represents that after removing SVM, only GAN is used for image generation and discrimination, and the performance deteriorates significantly, indicating that the generator model alone has poor effect in classification tasks, as shown in Figure 4.

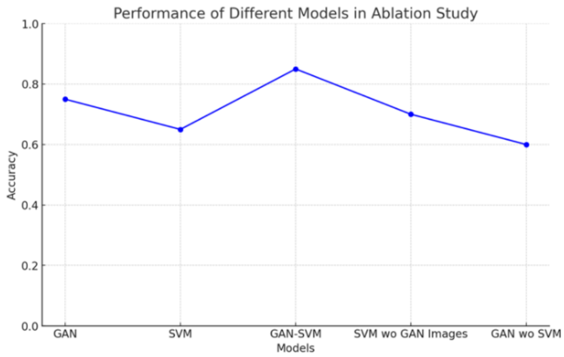

Figure 4. GAN-SVM ablation experiments were performed

The results of ablation experiments showed that the model's efficacy saw a substantial downturn following the detachment of GAN or SVM components, further demonstrating the importance of these two components in combined models. Specifically, when Gan (generating adversarial networks) are removed, the model's ability to generate high-quality images is significantly weakened, resulting in a decline in overall image quality. As a powerful generative model, Gan can effectively capture the complex distribution of data to generate realistic images, which plays a key role in improving model performance. On the other hand, after the removal of SVM (support vector machine), the classification accuracy and generalization ability of the model are significantly reduced. SVM, as a widely used classifier, is especially good at dealing with high-dimensional data and linear indivisible problems, and has excellent performance in constructing classification decision boundaries. Therefore, the introduction of SVM greatly enhances the model's efficacy in distinguishing tasks.

The results of these ablation experiments fully demonstrate the indispensability of GAN and SVM in the model, and their synergistic effect significantly enhances the holistic efficiency of the model. By combining the generation capability of GAN and the classification capability of SVM, the model can not only generate high-quality images, but also achieve excellent performance in classification tasks. This finding further validates the effectiveness of combining these two powerful components and provides a valuable reference for future model optimization.

4. Experimental summary

This study explores an innovative image recognition framework that integrates the powerful feature learning potential of generative adversarial networks with the stable classification performance of support vector machines (SVM), aiming to push the boundaries of image recognition technology. The research first reviewed the core advantage of GAN as a cutting-edge deep learning technology, that is, through the dynamic game between generator and discriminator, not only can create highly realistic images, but it also learns the core structural features of the dataset during this process, a trait highly beneficial for image recognition endeavors. At the same time, SVM's solid status as a classical algorithm has been emphasized, and its ability to efficiently divide data and handle linear and nonlinear classification problems in high-dimensional space has made it the preferred tool to strengthen image recognition systems.

In the experiment, we built a system that combines the two. First, by training the GAN model and using the real image data, the generator learns the real feature distribution of the simulated image, which not only generates new images, but also deepens the understanding of the image features. These real images, along with the resultant composite images, then form the SVM's training set, which not only enhances data diversity, but also provides the SVM with broader classification cues that help the model precisely locate classification boundaries in complex feature Spaces. To enhance the efficacy of the comprehensive framework, the study explored kernel selection and parameter adjustment in the SVM model, recognizing that these factors play a decisive role in the model's ability to handle nonlinear data and generalization performance. Through systematic evaluation and adjustment, we seek the most favorable configuration within the parameter domain to maximize the classification performance.

In summary, this study successfully demonstrated the potential of the image recognition framework combined with GAN and SVM, which not only demonstrated the complementary advantages of the two in theory, but also verified the remarkable effect of the framework in improving the efficiency and accuracy of image recognition through carefully designed experimental procedures in practice. Future work can further explore the application of more complex data sets, and how to fine-tune model parameters to adapt to different image recognition needs, further propelling advancements in the discipline.

References

[1] Xu, T., Li, I., Zhan, Q., Hu, Y., & Yang, H. (2024). Research on Intelligent System of Multimodal Deep Learning in Image Recognition. Journal of Computing and Electronic Information Management, 12(3), 79-83.

[2] Zhang, H., Diao, S., Yang, Y., Zhong, J., & Yan, Y. (2024). Multi-scale image recognition strategy based on convolutional neural network. Journal of Computing and Electronic Information Management, 12(3), 107-113.

[3] Wang, S., Liu, Z., & Peng, B. (2023, December). A Self-training Framework for Automated Medical Report Generation. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 16443-16449).

[4] Xiao, M., Li, Y., Yan, X., Gao, M., & Wang, W. (2024). Convolutional neural network classification of cancer cytopathology images: taking breast cancer as an example. arXiv preprint arXiv:2404.08279.

[5] Dai, W., Tao, J., Yan, X., Feng, Z., & Chen, J. (2023, November). Addressing Unintended Bias in Toxicity Detection: An LSTM and Attention-Based Approach. In 2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 375-379). IEEE.

[6] Mei, T., Zi, Y., Cheng, X., Gao, Z., Wang, Q., & Yang, H. (2024). Efficiency optimization of large-scale language models based on deep learning in natural language processing tasks. arXiv preprint arXiv:2405.11704.

[7] Lu, S., Liu, Z., Liu, T., & Zhou, W. (2023). Scaling-up medical vision-and-language representation learning with federated learning. Engineering Applications of Artificial Intelligence, 126, 107037.

[8] Liu, Z., & Song, J. (2021, November). Comparison of Tree-based Feature Selection Algorithms on Biological Omics Dataset. In Proceedings of the 5th International Conference on Advances in Artificial Intelligence (pp. 165-169).

[9] Zhao, B., Cao, Z., & Wang, S. (2017). Lung vessel segmentation based on random forests. Electronics Letters, 53(4), 220-222.

[10] Yan, X., Wang, W., Xiao, M., Li, Y., & Gao, M. (2024). Survival Prediction Across Diverse Cancer Types Using Neural Networks. arXiv preprint arXiv:2404.08713.

[11] Wang, Q., Schindler, S. E., Chen, G., Mckay, N. S., McCullough, A., Flores, S., ... & Benzinger, T. L. (2024). Investigating White Matter Neuroinflammation in Alzheimer Disease Using Diffusion-Based Neuroinflammation Imaging. Neurology, 102(4), e208013.

[12] Li, Y., Yan, X., Xiao, M., Wang, W., & Zhang, F. (2024). Investigation of Creating Accessibility Linked Data Based on Publicly Available Accessibility Datasets. In Proceedings of the 2023 13th International Conference on Communication and Network Security (pp. 77–81). Association for Computing Machinery.

[13] Yao, J., Wu, T., & Zhang, X. (2023). Improving depth gradient continuity in transformers: A comparative study on monocular depth estimation with cnn. arXiv preprint arXiv:2308.08333.