XLVI CILAMCE
VITÓRIA - ES - BRAZIL - 2025
XLVI IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING

Automated Detection of Exposed Rebar in Civil Structures Using UAV-Based Image Processing: Case Study in a Real Infrastructure

Vinícius dos Santos Mota[1], Elisa Dominguez Sotelino[1], Cristiano Saad Travassos do Carmo [1]

[1]*Departament of Civil and Environmental Engineering of Pontifical Catholic University of Rio de Janeiro (PUC-Rio)*
Rio de Janeiro, RJ, Brazil.
viniciusmota@aluno.puc-rio.br, sotelino@puc-rio.br, csaad@puc-rio.br

Abstract. This study proposes and evaluates deep learning models for the segmentation of exposed steel rebar in concrete structures — a critical damage type often linked to concrete delamination. The motivation emerged from technical discussions with the Municipal Infrastructure Secretaria(SMI) of Rio de Janeiro, which identified this pathology as a priority in the inspection of urban bridges. The dataset used includes approximately 100 images, combining real UAV-captured frames and publicly available databases, with precise binary masks manually annotated. Three segmentation architectures were evaluated: U-Net with ResNet-34 encoder, lightweight U²-NetP, and a hybrid model combining Swin Transformer and DeepLabV3. Training employed k-fold cross-validation, data augmentation, and grid search with BCE+Dice+Focal loss functions. Only U-Net and Swin+DeepLabV3 were tested in real UAV inspection footage. Results show that Swin+DeepLabV3 achieved higher accuracy in detecting rebar under diverse lighting and textures, outperforming U-Net especially in multi-instance scenarios. The study highlights the potential of modern semantic segmentation models for field-ready structural diagnostics and contributes to closing a gap in the literature regarding real-world applications focused on rebar exposure.

Keywords: Deep Learning; Exposed Rebar; Computer Vision; Unmanned Aerial Vehicles(UAVs); Structural Inspection

1 Introduction

Reinforced concrete structures in urban environments are subject to various types of deterioration over time. Among these, concrete cover delamination and exposed rebar stand out due to their direct impact on the structure's durability and safety. In cities with aging infrastructure, such as Rio de Janeiro, these pathological manifestations have become increasingly frequent. Technical consultations with the Municipal Infrastructure Secretariat (SMI) of Rio de Janeiro confirmed that concrete spalling and rebar exposure are among the most critical issues affecting the city's special civil structures, prompting the need for technological solutions to improve inspection and maintenance processes. Early detection of these anomalies is essential for ensuring public safety and optimizing maintenance strategies [1].

Recent advances in computer vision and deep learning have enabled the use of robust models for detecting and segmenting cracks, corrosion, and other structural damage directly from images. However, most studies focus on linear surface cracks, with comparatively little attention to damage involving rebar exposure, which presents more irregular and morphologically complex patterns, making accurate segmentation more challenging [2–4].

In this context, the present work proposes and evaluates different deep neural network architectures focused on detecting, locating, and segmenting areas with exposed steel rebar, aiming to bridge the gap between academic research and real-world engineering applications. Three segmentation architectures were implemented, trained, and validated: U-Net with ResNet34 encoder, U²-Net, and a hybrid model based on Swin Transformer combined with DeepLabV3. The models were compared using quantitative metrics (F1-score, IoU) and tested on real UAV-captured images of urban infrastructure.

As a key contribution, in addition to evaluation on annotated datasets, the study includes real-world field testing and a practical discussion of model performance in terms of accuracy, resilience to visual noise, and potential integration with digital maintenance workflows such as BIM/BrIM. These findings aim to support the development of AI-assisted visual inspection systems with direct applicability to municipal infrastructure management.

2 Related Works

The use of convolutional neural networks (CNNs) for civil infrastructure inspection has gained prominence in recent years, particularly for detecting fine cracks in concrete. Seminal studies by Dung & Anh [5], Cha et al. [6], and Liu et al. [7] demonstrated the effectiveness of architectures such as U-Net, FCN, and ResNet for crack segmentation in bridges and pavements under controlled conditions.

However, a broader review of recent literature shows that most research is still focused on surface cracks or generic damage types, with limited attention to the specific challenge of exposed rebar — a critical pathological manifestation, especially when linked to concrete delamination. Deng et al. [8] noted that visual interferences such as dirt or graffiti significantly hinder rebar detection, requiring more robust networks. Huang et al. [9] proposed a multi-class damage detection system for bridge inspection that includes rebar exposure, but with limited accuracy for this class.

While recent advances in multi-class segmentation have been reported (e.g., He et al. [10]; Santos and Carvalho [11]), there remains a shortage of models optimized specifically for the segmentation of exposed steel reinforcement — particularly using real UAV-acquired images with high-precision binary annotations, as explored in this work. A key challenge is the irregular morphology of exposed regions, which contrasts with the typically linear patterns of cracks [2].

Furthermore, few studies include real-world field validations. Although Jeong et al. [12] and Xiao et al. [13] used UAVs to collect data, their work still focuses on conventional cracks or 3D mapping. Regarding datasets, Kompanets et al. [14] and Deng et al. [15] emphasized the need for specific repositories for complex structural damage, including acquisition metadata and expert-validated annotations.

This study differentiates itself by directly addressing the precise segmentation of exposed rebar regions, using supervised training with custom binary masks. The proposed approach aims to fill a gap in the literature: the lack of field-ready solutions for delamination + rebar exposure damage, particularly using modern networks such as U²-Net and the Swin Transformer combined with DeepLabV3 — which remain underexplored for this application.

3 Methodology

This study was structured in three main stages: (i) dataset preparation and annotation, (ii) training and cross-validation of segmentation networks, and (iii) practical evaluation through a real-world case study using drone-captured imagery of a bridge with exposed reinforcement.

3.1 Dataset and Annotation

The dataset used for training and validation consisted of approximately 100 images, with around 60 showing exposed rebar (positive class) and 40 showing intact concrete surfaces (negative class). Image resolutions varied widely: negatives ranged from 227×227 to 256×256 pixels, while positives spanned 259×194 to 4608×3072 pixels, reflecting the diversity of real-world capture conditions.

Positive samples were sourced from the publicly available "Exposed-Steel-Rebar-Concrete" dataset on Kaggle, which includes images of exposed reinforcement under various lighting conditions and viewing angles. Negative samples came from two sources: the SDNET2018 database — widely used for crack detection in concrete — and six real images captured by UAV inspection of another bridge with surface features such as stains and graffiti, but no exposed rebar.

Manual annotation of positive images was performed using the Labelme tool, resulting in binary masks that precisely delineate exposed rebar regions. Despite being time-consuming, this step was essential for accurate

supervised training, providing clear spatial information on the target pathology.

Only models achieving a mean IoU above 0.60 and a mean F1-score above 0.70 across cross-validation folds were considered to have satisfactory performance and were thus selected for real-world inference using UAV inspection footage.

3.2 Training and Validation Strategy

Training followed a 5-fold cross-validation scheme to ensure robust evaluation and mitigate bias from limited data. For each network configuration, multiple combinations of learning rate and batch size were tested through grid search: learning rates of [1e-4, 5e-4, 1e-3] and batch sizes of [2, 4, 8].

The loss function used was a weighted combination of BCEWithLogitsLoss, Dice Loss, and Focal Loss, designed to balance penalization across small and large regions of exposed rebar, and to address pixel imbalance between background and target class. The Adam optimizer was employed, with adaptive learning rate reduction using the ReduceLROnPlateau scheduler.

During each fold, the model with the highest mean IoU was selected as the best for that fold. Early stopping was implemented with a patience of 10 epochs, and training was capped at 50 epochs. Images were normalized, and data augmentation was applied to the training set, including random horizontal flips, brightness/contrast adjustment, rotation, scaling, blur, and gamma transformations.

These augmentations were implemented using the Albumentations library. All images and masks were resized to 512×512 pixels during training. The training process was conducted on Google Colab Pro using an NVIDIA T4 GPU with 16GB of memory.

3.3 Evaluated Architectures

Three semantic segmentation architectures were tested: U-Net, U²-Net, and a hybrid model combining a Swin Transformer backbone with a DeepLabV3 decoder. These models were selected due to their relevance in the state of the art and complementary structural characteristics, which are critical in challenging scenarios such as UAV-based inspections — characterized by visual noise, lighting variability, limited labeled data, and scale diversity [2, 14].

U-Net, a well-established architecture originally designed for biomedical segmentation [17], has been widely adopted in structural inspection tasks. Its symmetric encoder-decoder structure with skip connections allows for the preservation of fine spatial details — crucial for detecting small pathologies on complex surfaces. Here, a ResNet-34 encoder pretrained on ImageNet was used, which showed robust performance under varying lighting and texture conditions, consistent with findings by Cha et al. [6] and Zhang et al. [16].

U²-Net [17] introduces nested Residual U-blocks, enabling efficient multi-scale contextual feature capture. This is especially useful for identifying small, partially occluded targets such as fine cracks or fragmented exposed rebar. The lightweight U²-NetP variant was used in this study to assess its computational feasibility on resource-constrained hardware, such as commercial UAVs.

Lastly, a hybrid architecture was tested using the Swin Transformer [18] for feature extraction, combined with a DeepLabV3 decoder. The Swin Transformer uses shifted window attention, providing efficient representation of irregular, multi-scale, and non-uniform visual patterns — typical of real inspection imagery. This model has shown strong performance in recent studies on UAV-based structural segmentation, including the work by Santos and Carvalho [11], who integrated it with the Segment Anything Model (SAM) and highlighted its potential for BIM, XAI, and multi-agent system integration.

3.4 Case Study: Arquiteto Paulo Casé Bridge

To evaluate the generalization of trained models in real inspection scenarios, a case study was conducted on the Arquiteto Paulo Casé Bridge, located over the Jardim de Alah canal in the southern zone of Rio de Janeiro, presented in figure 1. The structure spans approximately 20 meters in length and 17 meters in width, connecting the neighborhoods of Ipanema and Leblon. It plays a key role in local mobility and is considered a special structure from an engineering standpoint.



Figure 1. Arquiteto Paulo Casé Bridge

The bridge consists of reinforced concrete beams arranged as longitudinal girders and transverse beams. The girders have a cross-section of 1 meter height and 35 cm width, while the transverse beams measure 20 cm wide by 45 cm high — dimensions that reflect the structure's complexity and exposure to aggressive urban and marine environments.

Inspection was performed using a DJI Phantom 4 RTK SE drone, flying along the outer girder of one side of the bridge. Due to temporary fencing, it was not possible to inspect the opposite side, and access to the underside was limited by high water levels. The flight lasted approximately 1 minute and 40 seconds, capturing continuous video footage.

From this footage, 101 individual frames were extracted, all depicting the outer girder. Manual analysis revealed that 33 frames contained visible signs of exposed rebar, while the remaining 68 showed no evidence of the pathology. This dataset served as the inference testbed for the trained models, simulating a real-world UAV-based inspection scenario.

Importantly, these frames were not included in the training or validation sets, ensuring that the evaluation occurred on previously unseen data — a crucial condition for assessing practical generalization capability.

4 Results and Discussion

This section presents the results obtained by the tested segmentation architectures, beginning with training and cross-validation performance using the manually annotated dataset. The networks were evaluated based on standardized metrics such as Intersection over Union (IoU) and F1-score to determine which architecture best balanced accuracy and generalization before advancing to real-image evaluation.

4.1 Performance on annotated dataset

The U-Net architecture, combined with a ResNet-34 encoder pretrained on ImageNet, exhibited consistent and robust performance throughout training. The best configuration found via grid search employed a learning rate of 0.0001 and a batch size of 8, achieving a peak IoU of 0.6240 and an F1-score of 0.7692 in the top-performing fold. In addition to strong quantitative results, U-Net demonstrated high sensitivity in detecting elongated, narrow regions — typical of exposed rebar.

U²-Net, a deep encoder-decoder model featuring multi-scale residual blocks, underperformed relative to the other tested architectures. Its best configuration reached only an IoU of 0.3342 and an F1-score of 0.5006, despite hyperparameter tuning and use of the same normalization and loss strategies. While it benefited from faster training times and low computational demand, its generalization capability was insufficient for the task at hand.

The hybrid architecture combining a Swin-Tiny Transformer backbone with a DeepLabV3 decoder produced the best overall results on the validation set. Leveraging local tokens and hierarchical shifted windows, the Swin Transformer proved more robust in representing complex spatial regions. The optimal configuration (learning rate = 0.0001, batch size = 4) achieved an IoU of 0.6551 and an F1-score of 0.7902, outperforming all other models in both accuracy and inter-fold consistency.

Given these findings, only the U-Net (ResNet34) and Swin+DeepLabV3 models met the performance threshold for further analysis in real UAV footage. These models demonstrated superior and consistent validation results, clearly outperforming U²-Net in every evaluated metric.

The three architectures evaluated(U²-Net (Lite), U-Net (ResNet-34), and Swin Transformer + DeepLabV3) ,demonstrated varied performance levels on the annotated dataset. U²-Net consistently underperformed, yielding a mean IoU of only 0.3342 and an F1-score of 0.5006, which indicates limited suitability for the task. In contrast, U-Net achieved an IoU of 0.6240 and an F1-score of 0.7692, while Swin+DeepLabV3 reached an IoU of 0.6551 and an F1-score of 0.7902. These values, although slightly lower than benchmarks reported for crack segmentation tasks, which range from 0.70 to 0.82 in controlled conditions (Liu et al.[7], Zhang et al.[16]) , remain within acceptable bounds given the greater complexity of detecting exposed rebar. Unlike narrow and high-contrast cracks, rebar exposure often presents ambiguous textures, irregular geometries, and low contrast with surrounding materials, especially in UAV-captured imagery under uncontrolled lighting.

Given these findings, only the U-Net (ResNet-34) and Swin+DeepLabV3 models met the predefined performance threshold of mean IoU > 0.60 and mean F1-score > 0.70 across validation folds. These models exhibited consistent and robust behavior and were therefore selected for further testing on UAV footage. U²-Net, having failed to meet these thresholds, was excluded from the real-world inference stage.

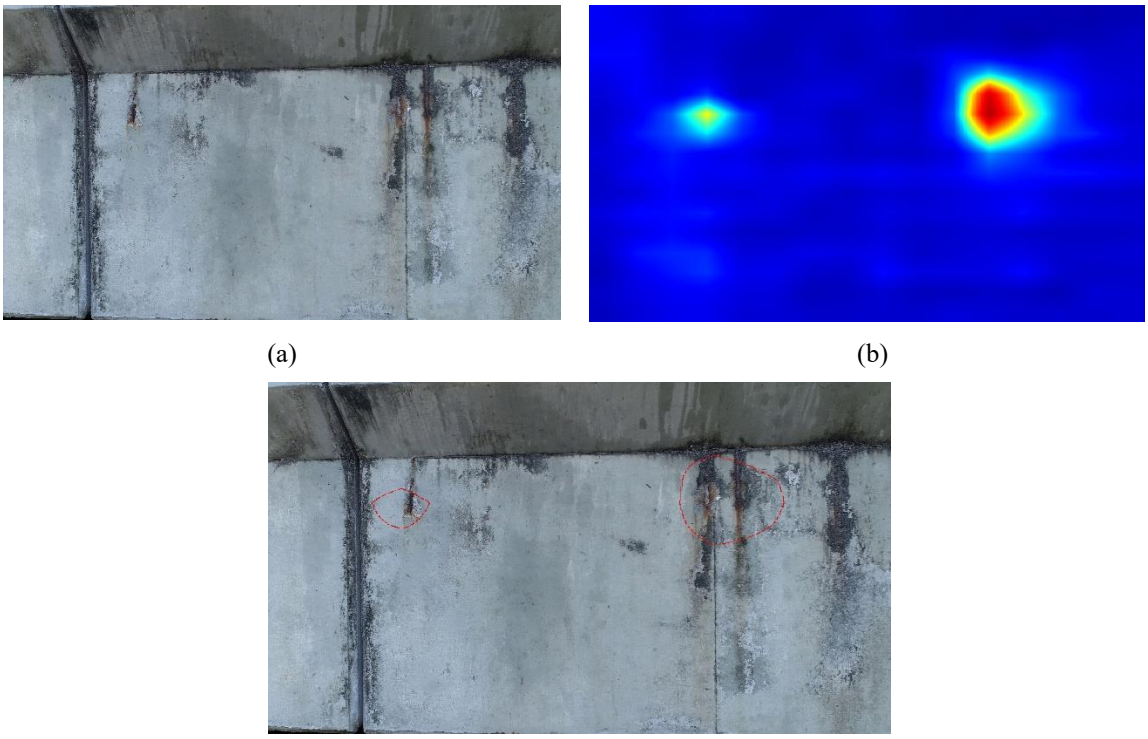
4.2 Case Study Result's

In the case study phase, the best models for each architecture — as determined by cross-validation — were used to process 101 frames extracted from UAV footage of the target bridge. For each image, a probability mask was generated and converted into a heatmap (blue to red scale) to enhance visual interpretation of potential rebar exposure. This output was then binarized using adaptive thresholding, and the resulting contours were overlaid on the original images to facilitate region identification.

The Swin Transformer model delivered substantially better performance than U-Net. Among the 33 frames with confirmed exposed rebar, U-Net correctly identified 25, corresponding to a detection rate of approximately 76%. However, it also generated 8 false positives from the 68 negative frames, mistakenly classifying dark areas such as expansion joints or surface stains as exposed rebar.

Conversely, the Swin Transformer-based model achieved a 94% detection rate, correctly identifying 31 of the 33 positive frames. It also showed a false positive rate of just 3%, misclassifying only 2 of the 68 negative frames. Notably, it was able to detect multiple rebar exposures within a single frame — a critical capability in complex field scenarios. An illustration of this output is presented in Figure 2, where subfigure 2a shows the original UAV frame, 2b displays the predicted probability heatmap (with blue indicating low probability of exposed rebar and red indicating high probability), and 2c presents the final segmented result with overlaid contours of the detected regions. These findings reinforce the robustness of transformer-based architectures for semantic segmentation in challenging inspection contexts using UAV-acquired imagery.

A comparative summary of the both evaluated architectures is presented in Table 1, including metrics such as IoU, F1-score, detection rate, and number of false positives — further highlighting the superior performance of the Swin+DeepLabV3 model under real-world conditions.



(a) (b)

(c)

Figure 2. Original UAV inspection frame (a), predicted probability mask (b), and segmented output with identified rebar locations (c).

Table 1. Comparative summary of the two architectures tested in the case study

Architecture	IoU	F1-Score	Detection Rate (Case Study)	False Negatives	False Positives
U-Net (ResNet-34)	0.6240	0.7692	76%	8	8
Swin+DeepLabV3	0.6551	0.7902	94%	2	2

5 Conclusions

This study investigated the application of deep neural networks for the segmentation of exposed rebar on concrete surfaces using UAV-acquired imagery. Three architectures — U-Net, U²-Net, and Swin Transformer combined with DeepLabV3 — were evaluated on a mixed dataset composed of real-world and publicly available images to assess their effectiveness in diverse inspection scenarios.

Among the tested models, U²-Net showed limited generalization capacity and was excluded from the real-world evaluation stage. The Swin Transformer architecture outperformed the others, achieving a 94% detection rate and a false positive rate of only 3% in the case study on the Arquiteto Paulo Casé Bridge, demonstrating high potential under challenging visual conditions.

Key contributions of this work include the use of transformer-based segmentation in real UAV inspections, a novel dataset annotated with high precision, and validation in an operational infrastructure asset — helping to close the gap between academic research and field application. However, the limited dataset size — particularly the small number of real negative samples — is acknowledged as a constraint that may affect model generalization. Future work may include expanding the dataset across different structures, urban contexts, and lighting conditions to improve model robustness. In addition, statistical tests could be incorporated to reinforce the validity of model comparisons.

Post-processing strategies and hybrid approaches combining classical computer vision with deep learning will also be explored to reduce false positives, especially in visually ambiguous regions such as stains or shadowed

joints.

Regarding practical adoption, a proposed integration pathway with BIM and BrIM platforms involves exporting segmented outputs as georeferenced layers or structured JSON files, allowing automatic updates of digital models based on field inspections. This can enable smart maintenance scheduling and historical tracking of structural deterioration.

In conclusion, while deep learning-based UAV inspection of civil structures shows high potential, its successful deployment requires not only accurate models but also lightweight and standardized tools for in-field operation. The convergence of UAV sensing, transformer-based segmentation, and BIM/BrIM-integrated workflows represents a promising direction toward scalable, real-time, and automated infrastructure diagnostics.

References

[1] Ibrahim A, Abdelkhalek S, Zayed T, Qureshi A, Mohammed Abdelkader EA. A comprehensive review of the key deterioration factors of concrete bridge decks. Buildings. 2024;14:1–38.

[2] Zhou, S., Canchila, C., & Song, W. (2023). Deep learning-based crack segmentation for civil infrastructure: data types, architectures, and benchmarked performance. Automation in Construction, 146. https://doi.org/10.1016/j.autcon.2022.104678

[3] Deng, J., Lu, Y., & Lee, V. C.-S. (2021). Imaging-based crack detection on concrete surfaces using You Only Look Once network. Structural Health Monitoring, 20(2), 484–499. https://doi.org/10.1177/1475921720938486

[4] Bae, H., Jang, K., & An, Y.-K. (2021). Deep super resolution crack network (SrcNet) for improving computer vision–based automated crack detectability in in situ bridges. Structural Health Monitoring, 20(4), 1428–1442. https://doi.org/10.1177/1475921720917227

[5] Dung, C. V, & Anh, L. D. (2019). Autonomous concrete crack detection using deep fully convolutional neural network. Automation in Construction, 99, 52–58. https://doi.org/10.1016/j.autcon.2018.11.028

[6] Cha Y-J, Ali R, Lewis J, Büyüköztürk O. Deep learning-based structural health monitoring. Automation in Construction. 2024;161:105328.

[7] Liu, Z., Cao, Y., Wang, Y., & Wang, W. (2019). Computer vision-based concrete crack detection using U-net fully convolutional networks. Automation in Construction, 104, 129–139. https://doi.org/10.1016/j.autcon.2019.04.005

[8] Deng, J., Lu, Y., & Lee, V. C.-S. (2020). Concrete crack detection with handwriting script interferences using faster region-based convolutional neural network. Computer-Aided Civil and Infrastructure Engineering, 35(4), 373–388. https://doi.org/10.1111/mice.12497

[9] Huang, L., Fan, G., Li, J., & Hao, H. (2024). Deep learning for automated multiclass surface damage detection in bridge inspections. Automation in Construction, 166. https://doi.org/10.1016/j.autcon.2024.105601

[10] He, Z., Su, C., & Deng, Y. (2024). A novel MO-YOLOv4 for segmentation of multi-class bridge damages. ADVANCED ENGINEERING INFORMATICS, 62(A). https://doi.org/10.1016/j.aei.2024.102586

[11] Santos, P. A. S., & Carvalho, M. T. M. (2025). Multi-class segmentation of structural damage and pathological manifestations using YOLOv8 and Segment Anything Model. Automation in Construction, 172. https://doi.org/10.1016/j.autcon.2025.106037

[12] Jeong, E., Seo, J., & Wacker, P. E. J. (2022). UAV-aided bridge inspection protocol through machine learning with improved visibility images. Expert Systems with Applications, 197. https://doi.org/10.1016/j.eswa.2022.116791

[13] Xiao, J.-L., Fan, J.-S., Liu, Y.-F., Li, B.-L., & Nie, J.-G. (2024). Region of interest (ROI) extraction and crack detection for UAV-based bridge inspection using point cloud segmentation and 3D-to-2D projection. Automation in Construction, 158. https://doi.org/10.1016/j.autcon.2023.105226

[14] Kompanets, A., Duits, R., Pai, G., Leonetti, D., & Snijder, H. H. B. (2025). Loss function inversion for improved crack segmentation in steel bridges using a CNN framework. Automation in Construction, 170. https://doi.org/10.1016/j.autcon.2024.105896

[15] Deng, L., Yuan, H., Long, L., Chun, P.-J., Chen, W., & Chu, H. (2024). Cascade refinement extraction network with active boundary loss for segmentation of concrete cracks from high-resolution images. Automation in Construction, 162. https://doi.org/10.1016/j.autcon.2024.105410

[16] Zhang, J., Qian, S., & Tan, C. (2022). Automated bridge surface crack detection and segmentation using computer vision-based deep learning model. Engineering Applications of Artificial Intelligence, 115. https://doi.org/10.1016/j.engappai.2022.105225

[17] Q. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand, "U²-Net: Going deeper with nested U-structure for salient object detection," *Pattern Recognition*, vol. 106, pp. 107404, 2020. DOI: 10.1016/j.patcog.2020.107404.

[18] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical vision transformer using shifted windows," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, vol. 2021, pp. 10012–10022, 2021. DOI: 10.1109/ICCV48922.2021.00986.