2024 11th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA) | 979-8-3315-2031-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICAICTA63815.2024.10763295

# Object Removal System for Urban Imagery Using Image Segmentation and Inpainting with A Deep Learning Approach

Eiffel Aqila Amarendra
School of Electrical Engineering and Informatics
Bandung Institute of Technology
Bandung, Indonesia
eiffelaqila@gmail.com

Rinaldi Munir
School of Electrical Engineering and Informatics
Bandung Institute of Technology
Bandung, Indonesia
rinaldi@staff.stei.itb.ac.id

***Abstract*—Digital imagery is a two-dimensional visual representation, which often holds emotional significance and crucial information. However, in images, specifically urban imagery, unwanted objects frequently appear. To address this issue, a system capable of automatically selecting areas with unwanted objects, removing these areas, and reconstructing the removed regions is essential.**

**The object removal system is developed by implementing and integrating an image segmentation module, image inpainting module, and graphical user interface application. The pre-trained image segmentation model, DeepLabv3+, is used for the image segmentation module. On the other hand, there are seven pre-trained image inpainting models, including DeepFillv2, EdgeConnect (Places), EdgeConnect (PSV), MADF (Places), MADF (PSV), MAT, and CoModGAN, which are compared across several testing aspects to be used in the image inpainting module.**

**Based on the analysis of the test results on the test data, the DeepLabv3+ model is proven to perform accurate segmentation with a mIoU value reaching 0.936. The CoModGAN model is chosen as the pre-trained model of the image inpainting module due to its average PSNR score of 26.59dB, SSIM of 0.8908, FID of 39.99, and subjective evaluation of 4.105. The graphical user interface application developed and integrated with the image segmentation and image inpainting modules successfully provides flexibility to users and shows increased performance compared to previous studies.**

***Keywords*—*object removal; urban imagery; image segmentation; image inpainting; deep learning***

## I. INTRODUCTION

Digital imagery is a two-dimensional visual representation, which often holds emotional significance and crucial information. However, in images, specifically urban imagery, unwanted objects frequently appear. The presence of undesired objects in urban imagery may present visual and aesthetical disturbances. Unwanted objects in this context may include pedestrians, cars, motorbikes, bicycles, trucks, buses, or other objects. Without the appearance of the unwanted object, the urban imagery is usable for background restoration, urban mapping, urban landscape reconstruction, and individual privacy protection. To address this issue, techniques that can restore the original appearance of the image by removing unwanted objects are essential, like in Fig 1 [1].

The image inpainting technique is an answer to overcome this problem. This technique can remove image areas with unwanted objects and fill them using cohesive information appropriate to the surrounding contexts so that the objects appear to have never existed [2]. Several studies have researched object removal using the image inpainting method. [3] and [4] has used the conventional image inpainting method. However, these two studies were still unable to produce accurate, rational, and aesthetic images of the inpainting results. Apart from that, the weakness in these two studies is the masking process for the undesirable areas is still done manually and randomly.

Along with the development of deep learning, inpainting techniques are starting to experience rapid progress. Deep learning-based inpainting provides inpainting results with better quality, accuracy, and rationality than traditional [2], [5]. Several studies have researched deep learning-based inpainting. [6] have utilized this method to restore damaged Persian pottery images. However, this study only focused on Persian pottery and did the mask manually. Besides that, in recent years, several deep learning-based inpainting models have been developed, such as DeepFill v2 [7], EdgeConnect [8], MADF [9], MAT [10], and CoModGAN [11]. These models have demonstrated excellent performance in image restoration and object removal.

To overcome the limitations of previous studies, this study proposed to integrate the image inpainting and image segmentation methods. By combining these methods, the potential for producing efficient solutions for removing unwanted objects in urban imagery increases. Furthermore, image inpainting can infer and fill the removed areas with known information appropriate to the surrounding contexts. [1] have integrated these methods and also focused on the urban scope. However, this study only has limitations on the object being restored, namely the building façades. Another limitation of this study is the inability to segment the shadows.

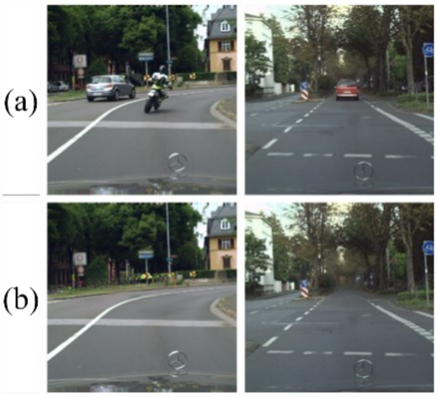

Fig. 1. The object removal results in urban imagery. (a) before removal (b) after removal.

979-8-3315-2031-1/24/$31.00 ©2024 IEEE

To address the limitations of previous studies, a more reliable object removal system specifically for urban imagery is proposed. The proposed system consists of an image segmentation module to provide accurate masking of unwanted objects, an image inpainting module to fill the masked areas, and an application that offers features for modifying the selection area of the unwanted objects. The image segmentation and image inpainting modules of the proposed system use deep learning-based algorithms.

## II. Literature Review

### A. Urban Digital Imagery

Urban digital imagery is a visual representation of the urban environment. Urban imagery contains information and insights about the corresponding city. This information includes city infrastructures, buildings, roads, public spaces, public facilities, social and economic activities, and environments. Apart from physical aspects, urban imagery can cover complex urban digital landscapes, including structure, governance, and people's interactions with the urban environment [12].

On the other hand, urban areas are geographical areas in the form of built-up areas consisting of concentrated built structures. This concentration includes a denser population and functional activities than the surrounding areas [13]. An urban area or paths that connect city areas, such as roads, sidewalks, and train tracks, edges or metropolitan area boundaries, districts or areas that have the same characteristics, strategic nodes or focal points, such as road intersections, and landmarks or physical objects to identify urban landscapes [14].

### B. Deep Learning

Deep learning is a concept that imitates the way of human brain works by using neural networks with a deep structure. Deep learning models are developed using artificial neural networks (ANN) [2]. An artificial neural network consists of several layers of neurons that are interconnected and weighted. Neurons function as message senders in the form of number signals that determine the weight of the corresponding signal, including the input layer that is responsible for receiving raw data, the intermediate/hidden layer that is responsible for extracting complex features, and the output layer that is responsible for generating conclusions, which is a prediction or classification [15].

The capabilities of deep learning trigger applications in various fields, such as computer vision, medical, and automotive. However, despite its capabilities, deep learning requires relatively massive resources, including large datasets and advanced computing devices [16]. Therefore, various studies have been developed to improve the efficiency and effectiveness of the deep learning method itself.

### C. Convolutional Neural Networks (CNN) for Object Removal System

Convolutional neural network (CNN) is one of the deep learning models. This model can learn spatial hierarchies between features in data automatically, adaptively, accurately, and concisely [17]. CNN commonly consists of several layers, including the convolution, pooling, and fully connected layers. Therefore, this model is often used in image processing, such as object detection, object recognition, and image segmentation.

First, the convolution layer is responsible for extracting features in the input image by applying convolution operations. Convolution is a mathematical operation that computes a dot product between the input data and a filter. This layer generally has an activation function at the end, namely ReLU, to prevent linearity on the network. Second, the pooling layer is responsible for down-sampling the feature maps result from the convolution layer. This process aims to decrease the complexity and size of the input so that it is easier to be processed in the following layers. Lastly, the fully connected layer is the last layer responsible for connecting every neuron and producing the final result. This layer is an artificial neural network that uses the softmax function to generate final predictions in the form of classes [18].

In recent years, CNNs have been increasingly applied in object removal systems, particularly in the domain of image segmentation and inpainting. [6] has utilized CNNs to restore damaged Persian pottery images, using partial convolution. Similarly, [1] has utilized CNNs to remove unwanted objects in facades, using DeepLabv3+ and DeepFill v2. This approach has demonstrated an excellent performance of CNNs in object restoration and removal. Furthermore, there are many recent image segmentation and inpainting models that have also used CNN algorithms to achieve high-quality results, particularly in urban imagery and other contexts.

### D. Generative Adversarial Networks (GAN) for Object Removal System

Generative adversarial networks (GAN) is a deep learning model that can produce realistic images. GAN uses a generative model to generate new data based on the training data. The model training process uses an opposition, namely an adversary. This model can manage diverse and massive datasets due to its deep learning approach [18]. This capability emerges in GAN applications in various image applications, such as image inpainting [2].

GAN uses two neural networks: a generator and a discriminator. The generator is the generative model that aims to manage training data distribution. The discriminator is the adversary that tries to identify the source of data, which is real or generated data. In the training process, these networks are competing. The generator focuses on improving its performance to produce more convincing results. On the other hand, the discriminator tries to improve its performance to distinguish between real and fake data [2], [19].

GANs have been extensively used in object removal systems due to their ability to generate highly realistic inpainting results. [1] has utilized GANs to remove unwanted objects in facades, using DeepFill v2. Similarly, many recent image inpainting models, such as EdgeConnect [8] and CoModGAN [11], also leverage GANs to achieve high-quality results. These models have demonstrated excellent performance of GANs in object removal by seamlessly removing objects and generating visually appealing results.

### E. Image Semantic Segmentation

Image segmentation is a process of partitioning a digital image into several segments at the pixel level. The image is segmented based on several characteristics, such as color, intensity, texture, and others. Semantic segmentation is an image segmentation that uses a deep learning approach to associate each image pixel with a certain label. This process combines several stages, including classification, detection, categorization, and object annotation for each pixel in the image. In contrast to instance segmentation which labels each object independently, semantic segmentation labels the same class for the same objects [20]. This image semantic segmentation's ability provides application opportunities in diverse fields, such as recognizing obstacles in facade images [1].

The labeling process in semantic segmentation is a major challenge due to the requirement of understanding the context and meaning of each image pixel. Apart from that, the main problems that arise in image segmentation can be divided into two, namely datasets and segmentation algorithms. One of the commonly used semantic segmentation algorithms is CNN due to its ability to extract features and produce accurate semantic segmentation results. There are several examples of models that utilize the CNN algorithm, including FCN, DeepLab, and Deconvnet CNN [20]

### F. *Image Inpainting*

Image inpainting is a technique for synthesizing context or restoring the empty, missing, or damaged areas of an image. This technique uses known information, such as structural and statistical information, to restore the missing image parts. Image inpainting is commonly used in various applications, such as object removal, image restoration, and image completion. In general, image inpainting methods can be divided into two, namely conventional and deep learning-based methods [21], [22].

The conventional image inpainting only uses simple features. There are several conventional image inpainting methods, such as the diffusion-based method that diffuses information to missing areas based on mathematical representation and the exemplar-based method that propagates known image exemplars to unknown areas. However, this approach has limitations in terms of image size and complexity [22].

On the other hand, deep learning-based image inpainting can predict the missing areas using the learned image semantic information. Using this information, this method can inpaint with better accuracy, quality, and rationality than the conventional approach. Several studies use a variety of deep learning models in image inpainting, such as CNN, GAN, or use them both using context encoder networks [2]. In the past few years, several deep learning-based inpainting models have been developed, such as DeepFill v2 [7], EdgeConnect [8], MADF [9], MAT [10], and CoModGAN [11].

## III. Proposed Solution

Fig 2 illustrates the overall workflow of the object removal system for urban imagery using image segmentation and inpainting with a deep learning approach. The proposed system consists of three main components, namely the image semantic segmentation module, image inpainting module, and graphical user interface (GUI) application. In general, the image segmentation module aims to automatically select unwanted objects. The image inpainting module aims to delete the segmented areas and reconstruct the removed areas. Meanwhile, the GUI application aims to receive the user's input parameters, display inpainted results, and provide various interactive features to users in the process of object removal. The object removal system source code was uploaded to Google Drive on the following page: https://drive.google.com/file/d/1HbFLUMytP3e-eViKY-omQ26DRGC2nady/view?usp=drive_link.

To provide a more detailed picture of data flow, Fig 3 illustrates the data flow diagram (DFD) level 0 of the proposed system. First, GUI receives three input parameters from the user, namely urban imagery with unwanted objects, the unwanted object selection, and pre-trained model selection. Next, the selected unwanted objects in the input image are segmented as a mask using the image segmentation module and a pre-trained model. Next, the image inpainting module removes the masked area of the input image and fills the removed area using the chosen pre-trained image inpainting model. Finally, the inpainted result is sent to the user through the GUI application.

### A. *Image Semantic Segmentation Module*

The image semantic segmentation module aims to segment the unwanted objects automatically from an image. The segmentation results from this module represent a mask that is used for the object removal process at the inpainting stage. This module is implemented using Python and a pre-trained image semantic segmentation model, DeepLabv3+, which is loaded using TensorFlow.

DeepLabv3+ is a CNN-based pre-trained image semantic segmentation model that applies encoder-decoder network architecture with atrous separable convolution, atrous separable pyramid pooling (ASPP), and Xception model as a backbone for the encoder network. This model uses the Cityscapes dataset and achieves an outstanding mIoU score of 82.1. In general, this model receives an image as an input and produces a colored segmentation image that distinguishes classes, including roads, sidewalks, buildings, walls, fences, poles, traffic lights, traffic signs, vegetation, terrain, sky, person, rider, car, truck, bus, train, motorcycle, bicycle, and void.

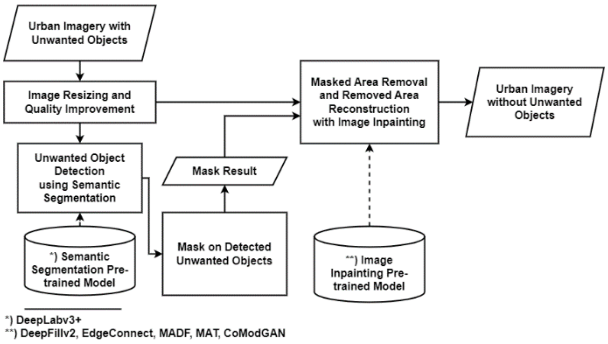
Urban Imagery with Unwanted Objects
Image Resizing and Quality Improvement
Unwanted Object Detection using Semantic Segmentation
*) Semantic Segmentation Pre-trained Model
Mask Result
Mask on Detected Unwanted Objects
Masked Area Removal and Removed Area Reconstruction with Image Inpainting
**) Image Inpainting Pre-trained Model
Urban Imagery without Unwanted Objects
*) DeepLabv3+
**) DeepFillv2, EdgeConnect, MADF, MAT, CoModGAN

Fig. 2. The overall workflow of the proposed system.

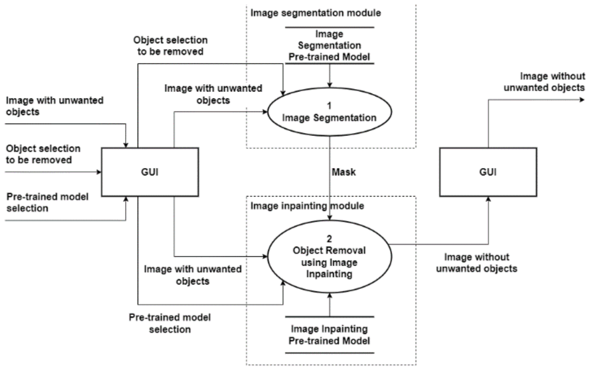
Image segmentation module
Image Segmentation Pre-trained Model
Object selection to be removed
Image with unwanted objects
1 Image Segmentation
Image without unwanted objects
Image with unwanted objects
Object selection to be removed
Pre-trained model selection
GUI
Mask
GUI
Image inpainting module
2 Object Removal using Image Inpainting
Image with unwanted objects
Image without unwanted objects
Pre-trained model selection
Image Inpainting Pre-trained Model

Fig. 3. The overall data flow of the proposed system.

Fig 4 shows the data flow of the image semantic segmentation module. This module receives images with unwanted objects and a selection of the unwanted objects. The selectable objects include pedestrians, riders, cars, motorcycles, trucks, buses, and trains. First, the module loads the input image and preprocesses it by resizing it to 512 × 512 pixels with the Lanczos interpolation technique. Next, the selected unwanted objects are segmented using the loaded pre-trained model. Then, the module postprocesses the segmented pixels by coloring, resizing, and expanding the image. The result of image coloring is a black-and-white mask that is done by changing selected unwanted objects into white, otherwise black. Image resizing aims to resize the mask results to the original image size. Lastly, the mask result is dilated so the unwanted objects are selected holistically.

### B. *Image Inpainting Module*

The image semantic segmentation module is the main component of the system. It aims to remove unwanted objects and synthesize context in the removed areas. The segmentation results from this module represent a mask that is used for the object removal process at the inpainting stage. This module is implemented using Python and pre-trained image inpainting models, namely DeepFillv2, EdgeConnect, Mask-Aware Dynamic Filtering (MADF), Mask-Aware Transformer (MAT), and CoModGAN, which is loaded using PyTorch and TensorFlow.

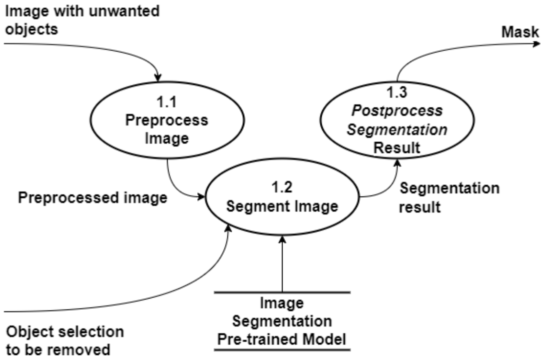
Image with unwanted objects
Mask
1.1
Preprocess Image
1.3
Postprocess Segmentation Result
Preprocessed image
1.2
Segment Image
Segmentation result
Image Segmentation Pre-trained Model
Object selection to be removed

Fig. 4. The data flow diagram of the image semantic segmentation module.

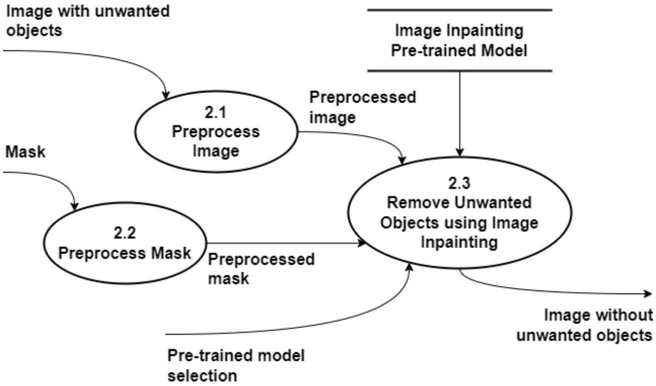
Image with unwanted objects
Image Inpainting Pre-trained Model
2.1
Preprocess Image
Preprocessed image
Mask
2.3
Remove Unwanted Objects using Image Inpainting
2.2
Preprocess Mask
Preprocessed mask
Image without unwanted objects
Pre-trained model selection

Fig. 5. The data flow diagram of the image inpainting module.

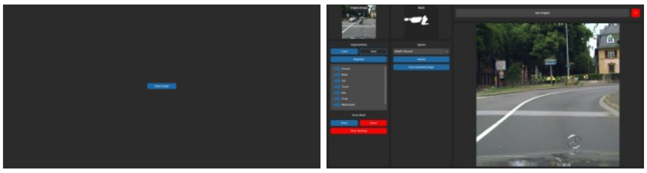

Fig. 6. The GUI application pages. Left: import image; Right: object removal.

The pre-trained image inpainting models use diverse architecture and training datasets. DeepFillv2 uses contextual attention and gated convolution in a two-stage coarse-to-fine network. EdgeConnect consists of two stages, namely edge generator and image completion network. MADF uses three main components, namely mask-aware dynamic filtering (MADF), point-wise normalization (PN), and end-to-end cascaded refinement. MAT uses a transformer with four main components, namely convolutional head, transformer body, convolutional tail, and style manipulation module. Lastly, CoModGAN applies co-modulation principles to generative adversarial networks (GAN). On the other hand, the datasets that are used for training these models, include Places for all of the models and Paris StreetView for EdgeConnect and MADF.

Fig 5 shows the data flow of the image inpainting module. This module receives images with unwanted objects and a mask. First, the module loads the input image and mask, then preprocesses it by resizing it to 512 × 512 pixels with the bilinear interpolation technique. The input image is also converted to RGB, while the mask is converted to grayscale. For certain cases, the mask is inverted to adjust the model requirements. Next, the unwanted objects are removed and filled with the appropriate contexts using the image inpainting technique. This process may differ between different pre-trained models. Lastly, the masked area of the image is removed and filled.

### C. *Graphical User Interface (GUI) Application*

The GUI application is the frontend component of the proposed system so that the user can interact. The motivation behind the application is to integrate the semantic segmentation and image inpainting module. Apart from that, this application aims to provide flexibility for the user to modify the segmentation result of the semantic segmentation module. This application is implemented as a desktop application using Python with Tkinter and CustomTkinter libraries. Fig 6 shows the implementation result of this GUI application.

This application provides several features, including import image, display input, modified mask, and inpainted images, automatic segment, and inpaint features. Besides that, the main feature of this application is the draw and erase mask feature for modifying the segmentation result or even building from scratch. Additionally, this application also provides additional features, such as choosing unwanted objects to be segmented, loading and saving a mask, choosing a pre-trained image inpainting model, and saving the inpainted result.

## IV. Testing and Evaluation

Object removal system testing aims to assess the system performance based on the system variable configuration, such as the pre-trained model selection and masking variation. The motivation for this testing is to determine the most optimal pre-trained model for the proposed system. The test results were uploaded to Google Drive on the following page: https://drive.google.com/drive/folders/1_rlYm8FvZN1tB7qLH-JQbaKtMEAf_2mn?usp=sharing.

### A. *Urban Imagery Test Dataset Making*

The urban imagery test dataset is a combination of open-source datasets, namely Paris StreetView and Cityscapes, and manually collected urban imagery from Indonesian cities. These datasets are manually selected to obtain appropriate data for the test dataset. This process filters out 140 image sets, consisting of urban imagery with unwanted objects and without unwanted objects (*ground truth*). However, since it is hard to obtain these image sets, some data are manipulated by placing unwanted objects in the urban imagery without unwanted objects. Lastly, this filtered test dataset is preprocessed to standardize the data by resizing it into pixels and adjusting the color to a reference image color distribution.

Besides the urban imagery test dataset, a mask dataset is also collected. This dataset consists of semantic segmentation results of images with unwanted objects. Next, this dataset is labeled to 4 mask ratios, including 1-10%, 10-20%, 20-30%, and 30-40%. From 1.886 of acquired data, 140 masks from each label are chosen randomly to fit the number of image sets. Lastly, this labeled dataset is preprocessed by resizing it into pixels and converting it into grayscale.

### *B. Image Segmentation Evaluation*

The image segmentation module testing measures the accuracy of the segmentation image using mean Intersection-over-Union (mIoU). Higher the mIoU means more accurate the model is There are 6 pairs of a segmentation result and a manually segmented image used in this testing. The test states that the DeepLabv3+ shows an outstanding performance with mIoU of 0.936 to the test data. It indicates that this model can segment the test data effectively and accurately. However, it still has problems when faced with CCTV and Indonesian cities' imagery.

### *C. Image Inpainting Evaluation*

The image inpainting module testing consists of two evaluations: qualitative and quantitative. The qualitative evaluation compares and analyses inpainting results from several models. There are 16 urban imagery sets used, including images with unwanted objects, masks with various ratios, and ground truth images. From the test, CoModGAN and MAT show that it can produce realistic and visually appealing inpainted results, without producing any blurs or checkerboard artifacts like the other models. However, these models sometimes produce random mismatched artifacts.

On the other hand, the quantitative evaluation compares models using several metrics, including PSNR, SSIM, and FID. The higher PSNR and SSIM mean the more accurate the model is, while the lower the FID means the more realistic the model is. Besides objective metrics, this testing also includes subjective evaluation through a survey to obtain insight from human visual perception directly. The higher the score given by the respondents means better the model is. There are 140 urban imagery sets used for the objective metrics and 16 urban imagery sets used for the subjective evaluation survey.

Fig 7 shows that mask ratio size significantly affects the performance of each model. In general, each model has tight average PSNR and SSIM scores with PSNR around 26 and SSIM around 0.89. It indicates that every model can successfully accurately inpaint the removed area. Even though PSNR and SSIM show the accuracy of the inpainting result to the ground truth image, they still can't show from a human visual perspective, since they only use pixel information individually. Therefore, it is crucial to use FID and subjective evaluation to obtain more comprehensive insights from the human visual perspective. Based on the FID and subjective evaluation results, CoModGAN shows impressive results compared to other models. It indicates that this model can generate a realistic, accurate, and visually appealing inpainting result with an average PSNR score of 26.59dB, SSIM 0.8908, FID 39.99, and subjective evaluation of 4.105. Therefore, CoModGAN is chosen as the pre-trained image inpainting model in the proposed system. Besides that, it also shows improved performance compared to previous studies, especially [1].

### *D. GUI Application Evaluation*

The GUI application testing evaluates the application sequentially to ensure the success of the implemented features and integration with the other components. There are 6 cases used in this test, including a successful case, a mask drawing case, a failed case, and other extreme cases. The test shows that the system can provide flexibility and perform segmentation and inpainting well. However, the system can't perform segmentation well in CCTV images, images with extreme lighting, and shadows. On the other hand, the inpainting process experiences decreased performance when faced with complex backgrounds and large sizes of unwanted objects.

### *E. Computation Efficiency Evaluation*

Table I shows that the average processing time of both modules is quite fast when not considering initialization time. The initialization time process can take almost 3 times longer because the pre-trained model loading process is very time-consuming. Apart from that, the evaluation results show that GPU RAM requirement is quite high so it requires sophisticated hardware. However, when both modules are run together, the GPU usage is lower, indicating efficiency in memory management.

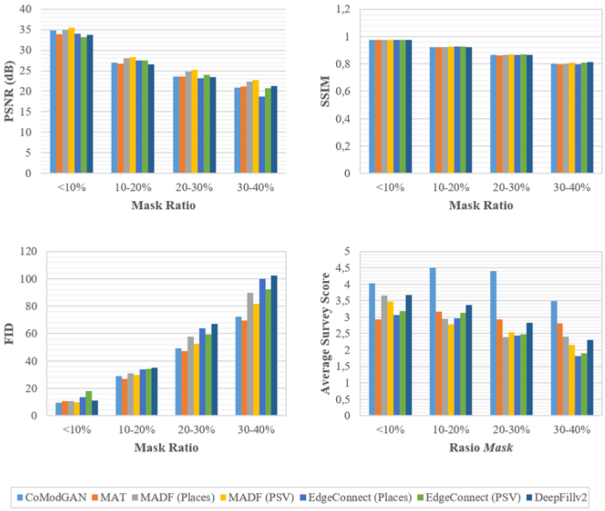

Fig. 7. The image inpainting quantitative evaluation results: PSNR (Upper-left; Higher, better), SSIM (Upper-right; Higher, better), FID (Lower-left; Lower, better), and subjective evaluation (Lower-right; Higher, better).

TABLE I. COMPUTATION EFFICIENCY EVALUATION RESULT

| **Module (Model)** | **Average Time (s)** | | **RAM Usage (GB)** |
|---|---|---|---|
| | ***With initialization time*** | ***Without initialization time*** | |
| Image semantic segmentation module (DeepLabv3+) | 16.445 | 10.048 | 3.376 |
| Image inpainting module (CoModGAN) | 16.028 | 2.485 | 2.57 |
| Total | 32.473 | 12.533 | 5.946 or 3.43[a] |

[a] RAM usage when both modules are run simultaneously

## V. Conclusion and Future Works

### A. Conclusion

The image segmentation module using the pre-trained model, DeepLabv3+, successfully performed segmentation with high accuracy, achieving mIoU of 0.936 on test data. However, it still has problems when faced with CCTV and Indonesian image data. The image inpainting module using the pre-trained model, CoModGAN, successfully performed inpainting with very high-quality results, namely achieving an average PSNR score of 26.59dB, SSIM 0.8908, FID 39.99, and subjective evaluation of 4.105. The integration of both modules into the GUI application of the object removal system successfully provides the user with the flexibility to modify the segmentation results. The developed object removal system shows improved performance compared with previous related studies, especially in terms of quality of inpainting results and flexibility of use.

### B. Future Works

For future works, the following suggestions can be implemented:

*1)* Develop a more sophisticated segmentation model for lighting variations, shadow presence, and camera angles, especially in CCTV imagery.

*2)* Improve inpainting performance on areas with complex textures, such as buildings with complex architecture and patterns, and large mask ratios, for example by integrating attention mechanism.

*3)* Optimize the system's processing speed for real-time applications, for example by using model compression and hardware acceleration techniques.

*4)* Expand the training dataset by adding more variations of urban images, especially from various cities in Indonesia, to increase the generalization ability of the model.

## Acknowledgment

The authors express gratitude to God for the blessings received throughout the completion of this research. The authors would like to acknowledge Mr. Dr. Ir. Rinaldi Munir, M.T. for guiding this research. The authors would also like to thank Tanoto Foundation for providing tuition assistance through scholarships for 3.5 years. Additionally, the researcher would like to thank their family, friends, and others for providing assistance and support throughout the research endeavor.

## References

[1] J. Zhang, T. Fukuda, and N. Yabuki, "Automatic object removal with obstructed facades completion using semantic segmentation and generative adversarial inpainting," *IEEE Access*, 2021, doi: 10.1109/ACCESS.2021.3106124.

[2] X. Zhang, D. Zhai, T. Li, Y. Zhou, and Y. Lin, "Image inpainting based on deep learning: A review," Feb. 01, 2023, *Elsevier B.V.* doi: 10.1016/j.inffus.2022.08.033.

[3] F. S. Manunggal, Liliana, and K. Gunadi, "Pembuatan Aplikasi Objek Removal dengan Menggunakan Exemplar-Based Inpainting," 2013.

[4] N. Zhang, H. Ji, L. Liu, and G. Wang, "Exemplar-based image inpainting using angle-aware patch matching," Dec. 01, 2019, *Springer International Publishing*. doi: 10.1186/s13640-019-0471-2.

[5] H. Xiang, Q. Zou, M. A. Nawaz, X. Huang, F. Zhang, and H. Yu, "Deep learning for image inpainting: A survey," *Pattern Recognit*, vol. 134, Feb. 2023, doi: 10.1016/j.patcog.2022.109046.

[6] N. Farajzadeh and M. Hashemzadeh, "A deep neural network based framework for restoring the damaged persian pottery via digital inpainting," *J Comput Sci*, vol. 56, Nov. 2021, doi: 10.1016/j.jocs.2021.101486.

[7] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, "Free-Form Image Inpainting with Gated Convolution," Jun. 2018, [Online]. Available: http://arxiv.org/abs/1806.03589

[8] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi, and M. Ebrahimi, "EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning," Jan. 2019, [Online]. Available: http://arxiv.org/abs/1901.00212

[9] M. Zhu *et al.*, "Image Inpainting by End-to-End Cascaded Refinement with Mask Awareness," Apr. 2021, doi: 10.1109/TIP.2021.3076310.

[10] W. Li, Z. Lin, K. Zhou, L. Qi, Y. Wang, and J. Jia, "MAT: Mask-Aware Transformer for Large Hole Image Inpainting," 2022. [Online]. Available: https://github.com/fenglinglwb/MAT.

[11] S. Zhao *et al.*, "LARGE SCALE IMAGE COMPLETION VIA CO-MODUL-ATED GENERATIVE ADVERSARIAL NETWORKS," 2021. [Online]. Available: https://github.com/zsyzzsoft/co-mod-gan.

[12] I. D. Imaykin, "DIGITAL IMAGE OF THE CITY, ON THE EXAMPLE OF SAKHALIN ISLAND," in *Problems of Preservation and Translation of Culture in Digital Age*, Institute for Peace and Conflict Research, 2024, pp. 36–41. doi: 10.31312/978-5-6048848-7-4-04.

[13] N. K. Pontoh and I. Kustiwan, *Pengantar Perencanaan Perkotaan*. Bandung: Penerbit ITB, 2009.

[14] K. Lynch, "The Image of the City," 1960.

[15] M. Islam, G. Chen, and S. Jin, "An Overview of Neural Network," *American Journal of Neural Networks and Applications*, vol. 5, no. 1, p. 7, 2019, doi: 10.11648/j.ajnna.20190501.12.

[16] J. H. Hwang, J. W. Seo, J. H. Kim, S. Park, Y. J. Kim, and K. G. Kim, "Comparison between Deep Learning and Conventional Machine Learning in Classifying Iliofemoral Deep Venous Thrombosis upon CT Venography," *Diagnostics*, vol. 12, no. 2, Feb. 2022, doi: 10.3390/diagnostics12020274.

[17] M. Lu and S. Niu, "A detection approach using lstm-cnn for object removal caused by exemplar-based image inpainting," *Electronics (Switzerland)*, vol. 9, no. 5, May 2020, doi: 10.3390/electronics9050858.

[18] A. Aggarwal, M. Mittal, and G. Battineni, "Generative adversarial network: An overview of theory and applications," Apr. 01, 2021, *Elsevier Ltd*. doi: 10.1016/j.jjimei.2020.100004.

[19] I. J. Goodfellow *et al.*, "Generative Adversarial Nets," 2014. [Online]. Available: http://www.github.com/goodfeli/adversarial

[20] F. Sultana, A. Sufian, and P. Dutta, "Evolution of Image Segmentation using Deep Convolutional Neural Network: A Survey," *Knowl Based Syst*, vol. 201–202, Aug. 2020, doi: 10.1016/j.knosys.2020.106062.

[21] S. Patil, V. S. Malemath, and S. Muddapur, "Enhanced Technique for Exemplar Based Image Inpainting Method," 2023, pp. 153–163. doi: 10.2991/978-94-6463-196-8_14.

[22] T. Shanmukhaprasanthi, S. M. Rayavarapu, Y. L. Lavanya, and G. S. Rao, "A Comprehensive Study of Image Inpainting Techniques with Algorithmic approach," in *2023 6th International Conference on Information Systems and Computer Networks, ISCON 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ISCON57294.2023.10112205.