Class-Specific Noise Injection for Improved Road Segmentation

Yukai Gu[1], Hao Shan[1], Penghui Ruan[1], and Yutong Gao[2,3(✉)]

[1] Beijing Jiaotong University, Beijing, China
[2] School of Information Engineering, Minzu University of China, Beijing, China
18112018@bjtu.edu.cn
[3] Key Laboratory of Trustworthy Distributed Computing and Service (MoE), Beijing University of Posts and Telecommunications, Beijing, China

Abstract. In this paper, we introduce a novel class-specific noise method designed for efficient data augmentation in the realm of road segmentation. This approach is rooted in the observation that in practical image segmentation, edges area of specific class often holds higher level of importance than interiors. Distinct from traditional data augmentation techniques, our method tailors the generation of noise based on the specific class. Through experimental validation, we demonstrate that our proposed approach can significantly bolster the mean intersection over union (miou) performance of models on test datasets. Our technique holds potential for a broad spectrum of image segmentation tasks, including but not limited to medical imaging and road segmentation.

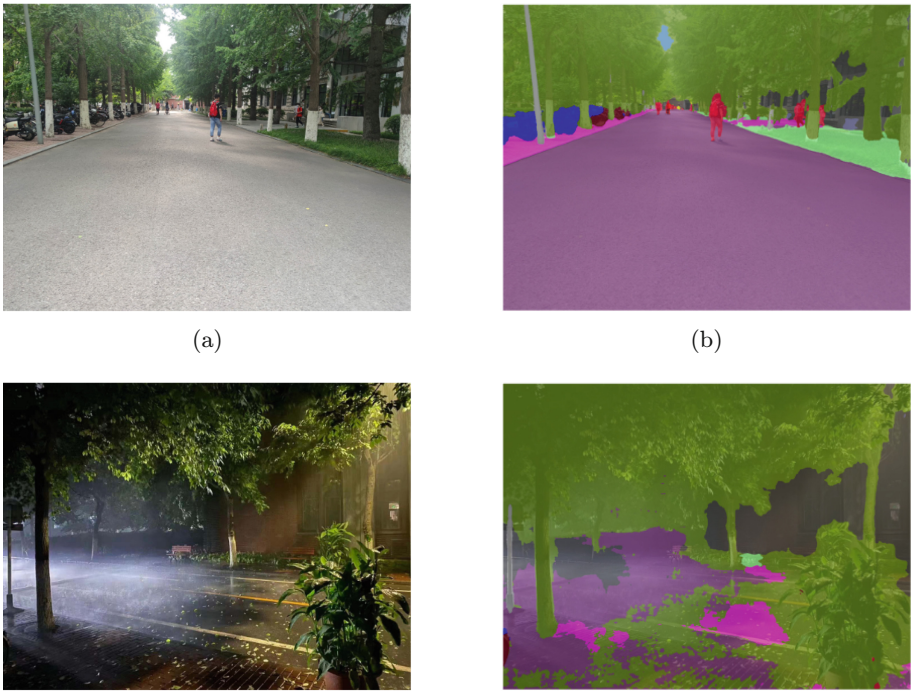
Keywords: Road segmentation · Image segmentation · Data augmentation · Computer Vision

1 Introduction

Image segmentation, a pivotal task in computer vision, seeks to divide an image into distinct segments or sets of pixels, each corresponding to different objects or features. This process is indispensable for a myriad of applications, spanning from medical imaging to the development of autonomous vehicles. Attaining a high degree of precision in segmentation necessitates not only sophisticated algorithms but also an extensive and varied training dataset. Unfortunately, the acquisition of such a comprehensive dataset is often a daunting and labor-intensive endeavor.

In the context of road segmentation for autonomous vehicles or traffic monitoring systems, the discrepancy between the ideal conditions present in training datasets and the multifaceted reality of road conditions is a significant challenge. Training datasets often feature roads that are clean, well-maintained, and uniformly lit by either natural or well-designed artificial lighting. However, the on-the-ground reality presents a far more complex and variable picture. As show in Fig. 1, (a) is sample of ideal road condition, while (b) is a sample of complex

W. Yu and L. Xuan (Eds.): MLICOM 2023, LNICST 518, pp. 96–106, 2024.
https://doi.org/10.1007/978-3-031-71716-1_8



(a) (b)

(c) (d)

Fig. 1. A demonstration of segmentation model. (a) (c) is the input image for segmentation, (b) (d) is the segmentation output of our baseline model.

road condition. We can see that the model act well in ideal condition, but it fails to segment the road in complex condition, misclassified the road as plant. In reality, countless factors can affect the quality of road images, including but not limited to weather, lighting, and occlusions. Consequently, models trained on pristine datasets may falter when confronted with real-world scenarios where road conditions are less than ideal, exhibiting varied textures and often littered with obstructions or wear.

Data augmentation, which entails artificially expanding the training dataset through modified versions of existing data, has emerged as an effective strategy to counteract the challenges of limited data. By introducing variability and extending the data manifold, data augmentation has consistently bolstered generalization, mitigated overfitting, and reinforced the robustness of deep learning models.

Although data augmentation is a staple in training deep learning models, its success often hinges on the quality of the augmentation strategy. Traditional techniques like random cropping, rotation, and flipping, though beneficial, may not always capture dataset-specific nuances. Take city image segmentation as an example: often, the noise in images is class-biased, bigger class like road in reality may particularly exposes to more occlusions that can't be fully covered in

image. Such nuances make traditional data augmentation methods less effective in enhancing model performance.

Recognizing that not all data augmentation methods are equally efficacious, we introduce a unique approach that selectively introduces noise to specific ground truth classes interiors. By adapting the augmentation process to the distinct characteristics of each class, our method fosters a deeper understanding of the dataset's intricacies, ultimately propelling improved segmentation performance, road segmentation in context of this work.

Our contributions are summarized as follows:

1. We propose a novel class-specific noise method for efficient data augmentation, which is proven to be efficient in complex road segmentation.
2. We proposed a new dataset for road segmentation featuring a higher prevalence of occlusions on roadways, which aims to serve as a benchmark for future research in high prevalence road segmentation.
3. We show that our method exhibits robust performance, enhancing the efficacy of various segmentation tasks across diverse conditions.

2 Related Work

We briefly review several aspects that are closely related to this paper, i.e., road segmentation, data augmentation and regularization.

2.1 Road Segmentation

Road Segmentation is generally considered as a semantic segmentation task, in which deep learning is widely applied. Most of these work concentrate on the adjusting of network structure. Existing regularization methods are not enough to tackle the problem of road segmentation in complex environment. [1,11,14,19,21,23,24,28] Some datasets are proposed for road segmentation too. [16,25]

Image segmentation has made notable strides with the adoption of deep learning techniques, and a pivotal innovation has been the application of transformer models. The introduction of Vision Transformer (ViT) encouraged a wave of its different variants. [3,18,26,27]

2.2 Data Augmentation

Data augmentation, which entails artificially expanding the training dataset through modified versions of existing data, has emerged as an effective strategy to counteract the challenges of limited data. We briefly review several aspects that are closely related to this paper [22].

Simple Transformations
Foundational image processing techniques often encompass flipping, rotation, cropping, and color jittering.

Flipping stands as one of the primal data augmentation methods, finding applicability and demonstrable utility in datasets like ImageNet and CIFAR-10. However, its application is context-sensitive; for instance, flipping can alter image labels, such as in the MNIST dataset, where a flipped "9" might resemble a "6".

Rotation, another cornerstone of data augmentation, is predominantly harnessed for image classification. Modest rotations typically augment model performance, allowing it to generalize better to varied orientations.

Cropping, which primarily involves extracting sub-regions from an image, is ubiquitously utilized across image processing tasks. Contrary to translation, cropping yields an image subset, typically smaller than the original.

Color jittering, though not elaborated here, involves random perturbations in the color space of the image, enhancing model robustness against varying lighting conditions and device-specific color encodings.

Noise Injection

Injecting noise into images serves as a strategy to ensure models are not overly sensitive to minute pixel-level variations. Techniques include Gaussian noise, salt-and-pepper noise, and speckle noise. For instance, medical imaging, which inherently contains noise due to equipment limitations, often benefits from noise injection as a form of data augmentation [17].

Color Space Transformations

Transformations in the color space involve modifying the color representation of an image. Techniques like histogram equalization, channel normalization, and conversion between RGB, HSV, or LAB color spaces can induce variability in the data. These transformations ensure models are robust against different lighting conditions, camera sensors, and post-processing effects [2,12].

Kernel Filters

Kernel-based filters, like Gaussian blur, sharpening, and edge-detection filters, alter image characteristics at a fundamental level. For instance, blurring can simulate out-of-focus scenarios or atmospheric distortions. Employing these filters as augmentation strategies can help models become resilient to varied image qualities and capture mechanisms [13].

Other Methods

Other methods use another network to generate optimal augmentation output. One way is to use neutral network to "smartly" learn an augmentation strategy [7,9,15,30].

Neural networks are powerful at mapping high-dimensional inputs into lower-dimensional representations. Some methods utilize this and directly augment on the feature space [8,29].

3 Methodology

Our methodology is rooted in the observation that, during the segmentation process, inner pixels of an object are often more prone to misclassification than its outer pixels. Consider road segmentation as an illustrative example: pixels situated in the middle of the road exhibit more variability than those on its edges. Standard datasets typically feature samples of roads under ideal conditions-clean, consistent in texture, and unmarred by real-world imperfections. Consequently, models trained on such datasets may falter when confronted with real-world scenarios where road conditions are less than ideal, exhibiting varied textures and often littered with obstructions or wear.

Addressing this challenge demands a targeted approach to data augmentation. Our strategy seeks to prompt the model to focus more on the edges of primary classes, which are regions often overlooked due to their consistent texture. Specifically, we inject noise into the central region of these primary classes, thereby ensuring the model does not become overly reliant on pristine and uniform textures during its learning process.

3.1 Class-Specific Noise Injection

Formulations. Let us consider a matrix $\boldsymbol{X}$ of size $H \times W$. Each pixel has been assigned to a class in $\{C_1, C_2, \cdots, C_i\}$. We denote by $\boldsymbol{Y}$ the ground truth matrix of size $H \times W \times$ where $\boldsymbol{Y}_{i,j} = c$ if the pixel (i, j) belongs to the class c and $\boldsymbol{Y}_{i,j,c} = 0$ otherwise.

We denote the noise matrix as $\boldsymbol{Z}$, which has dimensions $H \times W$. Prior to noise injection, it is essential to analyze the distribution of the ground truth area. We postulate that the class with the largest ground truth area is the most likely to suffer from insufficient generalization. Consequently, we introduce noise to this class, determined as:

$$C_{\text{chosen}} = \arg\max_{C_i} \text{area}\,(C_i)$$

where C_i represents each class in the dataset.

In our approach, we use Gaussian noise to the pixels within the chosen class C_{chosen} . The core idea is to modulate the intensity of noise in relation to the pixel's distance from the class boundary, ensuring that areas closer to the edge receive different noise levels compared to those deeper within the class. To achieve this, for every pixel p within C_{chosen} , we compute its distance $d(p)$ from the boundary of the class. The noise introduced to this pixel is then generated from a Gaussian distribution $N \sim \mathcal{N}\left(\mu, \sigma^2\right)$, scaled by the aforementioned distance. Mathematically, the noise for pixel p is given by Noise $(p) = d(p) \times N$. This noise value is then added to the original pixel intensity $\boldsymbol{X}(i, j)$ to yield the noise-injected pixel value $\hat{\boldsymbol{X}}(i, j) = \boldsymbol{X}(i, j)+$ Noise (p). By modulating noise injection in this manner, we ensure a spatially-sensitive disturbance, enabling a more nuanced augmentation of the dataset and potentially improving model robustness.

To encapsulate the entire process, the noise injection can be summarized as:

$$\hat{\boldsymbol{X}}(i,j) = \begin{cases} \boldsymbol{X}(i,j) + d(p) \times N & \text{if } \boldsymbol{Y}(i,j) = C_{\text{chosen}} \\ \boldsymbol{X}(i,j) & \text{otherwise} \end{cases}$$

Where:

- $\hat{\boldsymbol{X}}(i,j)$ represents the pixel value at position (i, j) in the noise-augmented image.
- $X(i,j)$ is the original pixel value at position (i, j) in the input matrix $\boldsymbol{X}$.
- $d(p)$ computes the distance of pixel p (which corresponds to position (i,j)) from the boundary of the class C_{chosen} which has the largest ground truth area.
- $N \sim \mathcal{N}\left(\mu, \sigma^2\right)$ is the Gaussian noise.
- $\boldsymbol{Y}(i,j)$ represents the class assignment of the pixel at position (i, j).
- $C_{\text{chosen}} = \arg\max_{C_i} \text{area}\,(C_i)$

3.2 Training Process

During training process, an initial step involves computing the class area for each sample and subsequently generating the noise-augmented object. Following this, the model is trained using the noise-perturbed image. It's crucial to highlight that only about 30% of the training samples are chosen from noise augmented sets to preserve the inherent information within the images. This ensures a balance between enhancing model robustness through exposure to perturbed data and retaining the original image characteristics essential for accurate feature learning. Too much noise could overwhelm the training process and obfuscate meaningful patterns in the data, while too little could render the noise augmentation process inconsequential. Therefore, selecting an optimal proportion of images for noise augmentation, such as the aforementioned 30%, is pivotal in achieving an effective trade-off between noise resilience and information fidelity.

3.3 Inference Process and High Occlusions Road Dataset

In Inference process, the network performs a forward process without any transformation applied to the images. One problem as stated before, is the lack of validation dataset. To solve this problem, we propose a new validation dataset composed of 200 images in city environment of various lighting, weather, occlusions, etc. The dataset images are collected from Internet and annotated by ourselves. We use this dataset to evaluate the performance of our model. The performance of our model on this dataset is shown in Sect. 4.2 (Fig. 2).

4 Experiments

In this section, we report image segmentation trained on CityScape [6] dataset, which includes 5000 2048×1024 color images for training and 1500 images for

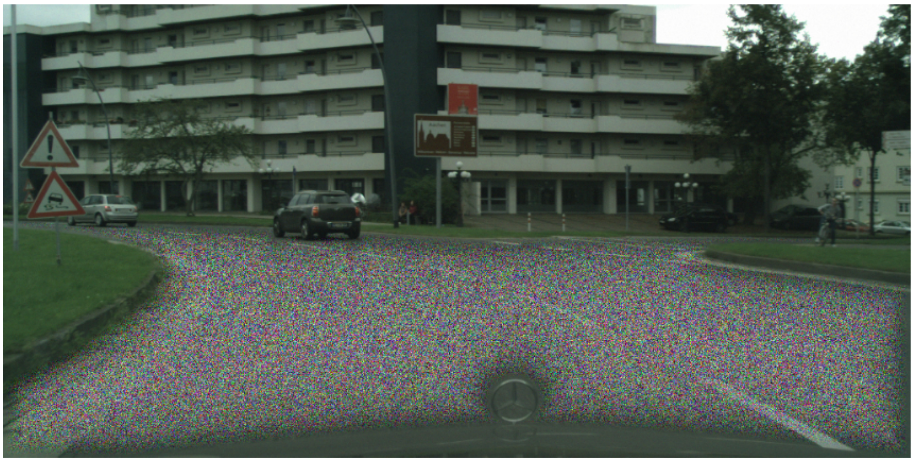


Fig. 2. A demonstration of augmented image

testing. The dataset contains 19 classes, including road, sidewalk, building, wall, fence, pole, traffic light, traffic sign, vegetation, terrain, sky, person, rider, car, truck, bus, train, motorcycle, bicycle. We show that our method can improve the performance of the model on the test set and its advantages over global noise injection.

4.1 Model Setting

In our experiment, we choose DeeplabV3 [4] as our baseline model. Note that some popular regularization techniques (i.e., weight decay, batch normalization and dropout) and various data augmentations (i.e., flipping, padding and cropping) are employed in the experiments. We use ResNet-101 as the backbone of the model. The model is trained on the CityScape dataset. We use the Adam optimizer with a learning rate of 0.0001. The batch size is set to 8. The model is trained for 100 epochs. We use the miou score as the evaluation metric. We compare our method with the baseline model and the model with global noise injection. For global noise injection, we add a Gaussian noise to the whole image. The noise is generated from a Gaussian distribution $N \sim \mathcal{N}\left(\mu, \sigma^2\right)$, where $\mu = 0$ and $\sigma = 5$. The noise is added to the image before the image is fed into the model. We also used grid mask presented in [5], mask ratio is set to 0.6.

We trained the model on the CityScape dataset and evaluated it on the validation set. The results are shown in Table 3. We trained with default setting, with global noise and with our method.

4.2 Baseline Performance on Different Road Dataset

After training the baseline model, we evaluate the IOU of road class on CityScapes, GTA [20], Road [10] and our own high occlusions road dataset.

As shown in Table 1, miou on road class on normal scenarios is high enough. To prove the capability of our method, we use our high occlusions dataset to evaluate the performance.

Table 1. Iou of Road Class of Baseline model On Different Dataset

Dataset	Overall Acc	Mean Acc	Freqw Acc	IOU
CityScape	95.9	83.7	92.3	95.8
GTA	87.5	64.4	79.9	93.1
Road	88.7	93.2	82.2	94.1
High Occlusions	85.1	81.0	78.9	73.8

4.3 Road Segmentation Performance on High Occlusions Dataset

These models are trained on Cityscape dataset but evaluated on our own dataset. As shown in Table 2, our method outperform the baseline model and the model with global noise injection.

Figure 3 shows the segmentation result of our method. We can see that our method can segment the road in high occlusions environment.

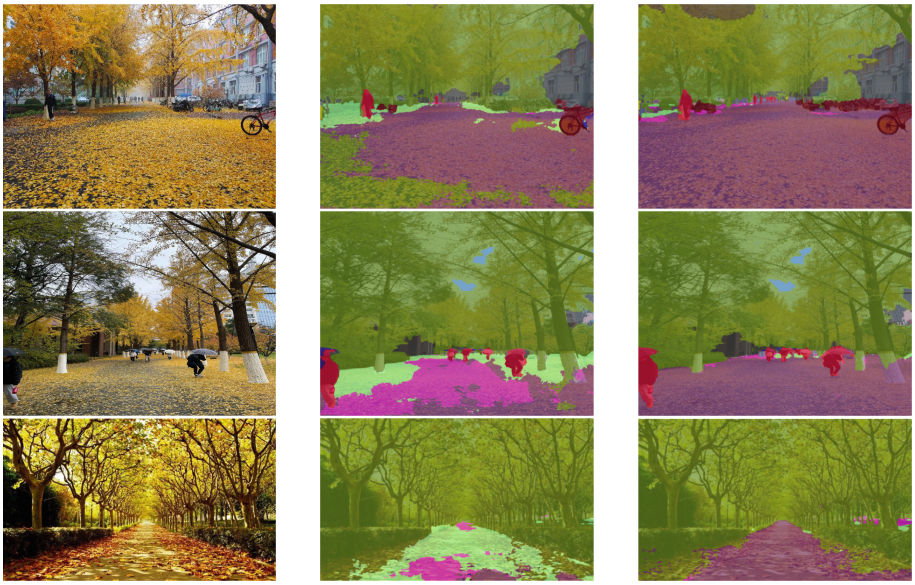
Table 2. Result miou

Model	miou
baseline	73.8
+ Global Noise	74.3
+ Grid Mask	73.8
+ Ours	**81.2**

4.4 Segmentation Performance On CityScape Val Set

As shown in Table 3, our method outperform the baseline model and the model with global noise injection. The miou score of our method is 77.5, which is 1.3 higher than the baseline model and 0.8 higher than the model with global noise injection.

Notice that the miou data is calculated on all classes. The result shows that our method can improve the performance of the model not only on road class, but also on other classes. This proves that our method can improve the robustness of the model.



(a) Input (b) Baseline (c) Ours

Fig. 3. The demonstration of our model on our own dataset. (a) is the input image, (b) is the segmentation output of baseline model, (c) is the segmentation output of our model.

Table 3. Result miou

Model	miou
baseline	76.2
+ Global Noise	76.5
+ Grid Mask	76.7
+ Ours	**77.5**

5 Conclusion

This paper introduces a new class-specific noise method for efficient data augmentation in regularization of road segmentation. This method is easy to implement and can be used in various image segmentation tasks, experiment proves its effectiveness in improving model's robustness to noise and occlusions. In the future, we will further explore the use of this method in other tasks, such as object detection, image classification, etc.

Acknowledgement. This work is supported by Major Projects of National Natural Science Foundation of China (Grant No. 72293583), Research on Privacy Data Protection and Iatrogenesis Decision of IHS.

References

1. Aksoy, E.E., Baci, S., Cavdar, S.: SalsaNet: fast road and vehicle segmentation in LiDAR point clouds for autonomous driving. In: 2020 IEEE Intelligent Vehicles Symposium (IV), pp. 926–932 (2020). https://doi.org/10.1109/IV47402.2020.9304694, https://ieeexplore.ieee.org/abstract/document/9304694, ISSN: 2642-7214
2. Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: delving deep into convolutional nets. arXiv preprint arXiv:1405.3531 (2014)
3. Chen, J., He, Y., Frey, E.C., Li, Y., Du, Y.: ViT-V-Net: vision transformer for unsupervised volumetric medical image registration. arXiv preprint arXiv:2104.06468 (2021)
4. Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation (2017)
5. Chen, P., Liu, S., Zhao, H., Jia, J.: GridMask data augmentation (2020)
6. Cordts, M., et al.: The cityscapes dataset for semantic urban scene understanding (2016)
7. Cubuk, E.D., Zoph, B., Mané, D., Vasudevan, V., Le, Q.V.: AutoAugment: learning augmentation policies from data. arXiv: abs/1805.09501 (2018). https://api.semanticscholar.org/CorpusID:43928340
8. DeVries, T., Taylor, G.W.: Dataset augmentation in feature space. arXiv preprint arXiv:1702.05538 (2017)
9. Duan, S., Zhao, H., Zhang, D.: Syntax-aware data augmentation for neural machine translation. IEEE/ACM Trans. Audio Speech Lang. Process. (2023)
10. Fritsch, J., Kuehnl, T., Geiger, A.: A new performance measure and evaluation benchmark for road detection algorithms. In: 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), pp. 1693–1700. IEEE (2013)
11. Jiqing, C., Depeng, W., Teng, L., Tian, L., Huabin, W.: All-weather road drivable area segmentation method based on CycleGAN. Vis. Comput. **39**(10), 5135–5151 (2023)
12. Jurio, A., Pagola, M., Galar, M., Lopez-Molina, C., Paternain, D.: A comparison study of different color spaces in clustering based image segmentation. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.) Information Processing and Management of Uncertainty in Knowledge-Based Systems. Applications: 13th International Conference, IPMU 2010, Dortmund, Germany, 28 June–2 July 2010, Proceedings, Part II 13, pp. 532–541. Springer, Cham (2010). https://doi.org/10.1007/978-3-642-14058-7_55
13. Kang, G., Dong, X., Zheng, L., Yang, Y.: PatchShuffle regularization. https://doi.org/10.48550/arXiv.1707.07103, http://arxiv.org/abs/1707.07103
14. Lan, M., Zhang, Y., Zhang, L., Du, B.: Global context based automatic road segmentation via dilated convolutional neural network. Inf. Sci. **535**, 156–171 (2020). https://doi.org/10.1016/j.ins.2020.05.062, https://www.sciencedirect.com/science/article/pii/S0020025520304862
15. Lemley, J., Bazrafkan, S., Corcoran, P.: Smart augmentation learning an optimal data augmentation strategy. IEEE Access **5**, 5858–5869 (2017)
16. Lu, J., Liu, H., Yao, Y., Tao, S., Tang, Z., Lu, J.: HSI road: a hyper spectral image dataset for road segmentation. In: 2020 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6. IEEE (2020)

17. Mikołajczyk, A., Grochowski, M.: Data augmentation for improving deep learning in image classification problem. In: 2018 International Interdisciplinary PhD Workshop (IIPhDW), pp. 117–122 (2018). https://doi.org/10.1109/IIPHDW.2018.8388338, https://ieeexplore.ieee.org/abstract/document/8388338
18. Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D.: Image segmentation using deep learning: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **44**(7), 3523–3542 (2021)
19. Oliveira, G.L., Burgard, W., Brox, T.: Efficient deep models for monocular road segmentation. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4885–4891. IEEE (2016)
20. Richter, S.R., Vineet, V., Roth, S., Koltun, V.: Playing for data: ground truth from computer games (2016)
21. Shamsolmoali, P., Zareapoor, M., Zhou, H., Wang, R., Yang, J.: Road segmentation for remote sensing images using adversarial spatial pyramid networks. IEEE Trans. Geosci. Remote Sens. **59**(6), 4673–4688 (2021). https://doi.org/10.1109/TGRS.2020.3016086, https://ieeexplore.ieee.org/abstract/document/9173823
22. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. J. Big Data **6**(1), 60 (2019). https://doi.org/10.1186/s40537-019-0197-0, https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0197-0
23. Sun, Z., Geng, H., Lu, Z., Scherer, R., Woźniak, M.: Review of road segmentation for SAR images. Remote Sens. **13**(5), 1011 (2021)
24. Valente, M., Stanciulescu, B.: Real-time method for general road segmentation. In: 2017 IEEE Intelligent Vehicles Symposium (IV), pp. 443–447. IEEE (2017)
25. Wang, H., et al.: SFNet-N: an improved SFNet algorithm for semantic segmentation of low-light autonomous driving road scenes. IEEE Trans. Intell. Transp. Syst. **23**(11), 21405–21417 (2022)
26. Wang, W., Tang, C., Wang, X., Zheng, B.: A ViT-based multiscale feature fusion approach for remote sensing image segmentation. IEEE Geosci. Remote Sens. Lett. **19**, 1–5 (2022)
27. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: SegFormer: simple and efficient design for semantic segmentation with transformers. Adv. Neural. Inf. Process. Syst. **34**, 12077–12090 (2021)
28. Zhang, L., Lan, M., Zhang, J., Tao, D.: Stagewise unsupervised domain adaptation with adversarial self-training for road segmentation of remote-sensing images. IEEE Trans. Geosci. Remote Sens. **60**, 1–13 (2022). https://doi.org/10.1109/TGRS.2021.3104032, https://ieeexplore.ieee.org/abstract/document/9516689
29. Zhang, Y., Zhu, H., Song, Z., Koniusz, P., King, I.: COSTA: covariance-preserving feature augmentation for graph contrastive learning. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 2524–2534 (2022)
30. Zhou, L.: Spatially exclusive pasting: a general data augmentation for the polyp segmentation. In: 2023 International Joint Conference on Neural Networks (IJCNN), pp. 01–07. IEEE (2023)