UMoncton CCNB-Innov Road Detection Preliminary Results

M-A. Blais, M. Akhloufi
Perception, Robotics and Intelligent Machines (PRIME),
Université de Moncton,
18 Antonine Maillet Ave.
Moncton, NB E1A 3E9

Introduction

Logging roads, whether active or inactive, have a profound impact on the planet and its fragile ecosystems. These roads are constructed, either temporarily or permanently, to facilitate the transportation of logs to main highways. Typically, they consist of compacted dirt, only a few meters wide, sometimes reinforced with gravel for stability. Additionally, these roads often include culverts and drainage pipes to divert or block water flow entirely. The scale of these networks is immense, stretching over hundreds of thousands of kilometers per province in Canada. For instance, Quebec alone has approximately 476,000 kilometers of such roads (Jutras, 2022). In New Brunswick, over 50,000 kilometers of roads are actively monitored, along with well over 10,000 kilometers of watercourse crossings (Natural Resources and Energy Development NB).+

Logging roads can severely affect ecosystems by contributing to erosion, increasing the risk of fires, triggering landslides, and causing sediment accumulation in water bodies (Kleinschroth, 2017). Aquatic environments, such as lakes, rivers, and wetlands, are particularly vulnerable to the disturbances caused by these roads (Jutras, 2022). In wetlands, logging roads are a significant source of sediment, which disrupts the delicate balance required for wetland function (Anderson, 1996). In New Brunswick, wetlands play a critical role in supporting local ecosystems and communities by providing essential services such as water filtration, sediment retention, and flood prevention. Despite these benefits, logging roads pose a significant threat to their integrity.

The environmental impact of logging roads can be mitigated through proper planning and the use of conventional engineering techniques (McCashion, 1983). However, it is only in recent years that the potential damage caused by these roads has been seriously considered in road network planning. As a result, many older road networks, both active and abandoned, remain unmapped and unmonitored. Mapping these roads is vital for researchers to assess their environmental impact and to inform the planning of future road networks. However, due to the sheer scale of these networks, manually mapping them is not feasible.

To address this challenge, we propose the use of artificial intelligence to automatically detect and map logging roads in satellite images. Specifically, our approach involves leveraging multiple segmentation algorithms to map the roads. Since this project is exploratory, we aim to conduct a broad evaluation of segmentation algorithms across four different image sizes. While we worked with a small training dataset comprising only a few images, the results were promising. Although our F1-Score did not surpass 0.60, the visual outputs show that most roads were successfully detected. We believe that increasing the size and quality of our dataset will significantly improve performance.

In the following sections, we present our proposed approach, detail the data used, and share our results. We conclude with a preliminary analysis and outline future directions for this project.

Related Works

As discussed in the previous section, we propose leveraging artificial intelligence to detect logging roads in satellite imagery. Currently, most existing approaches focus on detecting paved roads, such as highways (Ayala, 2021) (Mokhtarzade, 2007). These algorithms typically perform exceptionally well, largely due to the availability of extensive datasets, which provide models with a wealth of training images, enabling them to achieve very good results. In contrast, when it comes to detecting logging roads, several studies have explored artificial intelligence as a potential solution(Botelho Jr, 2022) (Sloan, 2024). While these studies show promising results, they often fall short in terms of precision. For example, the best models from these works achieved F1 scores of 68.4% and 81%, respectively, with (Sloan, 2024) additionally reporting a union score of 58%. These results highlight the challenges these models face in reaching high performance levels. The main limiting factors are typically the scarcity of training data and the quality of the available datasets.

When analyzing the visual output of these models, one common issue is the over-prediction of roads—meaning that the models tend to detect extra roads rather than miss existing ones. This can often be attributed to roads that are present in satellite images but absent in the segmentation masks. These unmapped roads may include farm lanes, also known as two-trackers or guide roads, which are primarily used by tractors for agricultural purposes. Due to their highly variable nature, changing from year to year, these roads are often not marked in the segmentation maps, further complicating the training process.

Proposed Approach

We propose comparing various algorithms and image sizes to detect logging roads in North-West New Brunswick. This study is not aimed at introducing new methods or results; rather, it serves as a feasibility evaluation of existing approaches. We utilize five categories of algorithms, encompassing a total of 22 individual models. The algorithms used in this evaluation are listed in Table 1. Additionally, we assess the performance of these algorithms across four different image crop sizes: 128, 256, 512, and 1,024 pixels.

Table I: Algorithms used

DenseNet	Inception	EfficientNet	ResNet	VGG
DenseNet121	Incepction-ResNetV2	EfficientNetB0	ResNet18	VGG11 (+bn)
DenseNet161	InceptionV4	EfficientNetB1	ResNet34	VGG13 (+bn)
DenseNet169	Xception		ResNet50	VGG16 (+bn)
DenseNet201			ResNet101	VGG19 (+bn)
			ResNet152	

We trained our models using the Adam optimizer with a learning rate of 0.00001 for 200 epochs. To evaluate the performance of the algorithms, we used the F1-score, which is a mean of precision and recall. Precision is defined as the accuracy of positive predictions, while recall measures the model's ability to identify all relevant instances. Equation 1 defines precision, followed by Equation 2 for recall, and Equation 3 for the F1-score.

$$Precision = \frac{True\ Positives}{True\ Positives + \ False\ Positives}$$

Equation 1

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

Equation 2

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Equation 3

Using the F1-score, we selected the best-performing model for each crop size and conducted a visual analysis of the results. Additionally, we compared the models with and without data augmentation to assess its impact on performance.

Data

Our dataset was provided by CCNB-Innov and was generated using ArcGIS, a modeling software. It was quickly modified to add or remove roads that were either present or missing in the images. The data provided was of moderate quality, with some inaccuracies in the masks, such as missing roads or extra roads. After a manual review, we selected nine image-mask pairs for training and four pairs for validation.

This dataset was further transformed into eight different datasets, representing four image sizes (128, 256, 512, and 1024 pixels) with both augmented and non-augmented versions. Each original image was randomly cropped to its respective size, with the condition that at least 3% of each class be present in the image to account for portions with missing masks.

For the augmentation datasets, we applied a series of transformations, including rotations, brightness adjustments, and Gaussian blur. In total, each dataset contained 900 training pairs and 400 validation pairs. It is important to note that the validation datasets, regardless of size, did not include any augmentation. This approach ensured a fair comparison between augmented and non-augmented models, as both were evaluated using the same validation data.

Results

In terms of F1-score, our models underperformed, with validation scores often stagnating at around 45%, while training scores exceeded 90%. This disparity suggests that the models struggled to generalize due to the limited number of images and the moderate quality of the dataset. Expanding the dataset size would likely alleviate this issue and improve performance.

Additionally, the augmented dataset for the largest crop size (1,024 pixels) did not converge, with the model frequently predicting the entire image as having no roads. Further investigation into this behavior will be necessary in future iterations.

Table 2 presents the validation F1-scores for each crop size, comparing models trained with and without data augmentation.

Table II: F1-Scores

128		**256**		**512**		**1,024**	
Aug	NoAug	Aug	NoAug	Aug	NoAug	Aug	NoAug
0.4659	0.4297	0.4920	0.4500	0.4920	0.4474	Error	0.4329

As shown in Table 2, the augmented datasets generally achieved slightly better F1-scores compared to their non-augmented counterparts. Despite these improvements, the results were not initially promising, as the models did not seem to have fully converged. However, a closer examination of the visual results reveals that the models produced promising initial outputs.

In Figure 1, we present the original image alongside its mask, as well as the model predictions for each crop size without augmentation. Since we used the original 5000x5000 image-mask size, we predicted the entire image by dividing it into non-overlapping cropped sections matching the training size. Due to the large image size, overlapping predictions occurred at the edges, but only the first predictions were included in the final stitched image.

This figure highlights how the original mask contained significant missing portions, while the predictions managed to capture roads not present in the mask. It also illustrates the reason for our models' low F1-scores: the predictions often contained substantial noise, likely due to the limited amount of training data. Increasing the dataset size would significantly reduce this noise, and incorporating post-processing techniques could further enhance the predictions.

Figure 2 provides a cropped view of the predictions, focusing on a 1,024x1,024 pixel section starting from the top-left corner. This figure more clearly demonstrates the relationship between the training image size and the noise in the predictions. We observed that larger crop sizes resulted in proportionally less noise, as the models were unable to pick up finer details, thus reducing the potential for noise in the predictions. However, this also led to a decrease in road detection accuracy, as smaller roads were often missed in larger images. This effect presents a trade-off, where reducing noise also reduces the ability to accurately detect smaller features.

Despite these challenges, the results are promising given that the models were trained on a small, low-quality dataset. Further training with a larger, higher-quality dataset could yield substantial improvements in performance.

Conclusion

In conclusion, logging roads significantly impact forests, wetlands, and the health of both local and global ecosystems. They contribute to increased erosion, fire risks, altered animal behavior, and various other environmental challenges. Effective planning and management of these roads are crucial for sustainable forest management and long-term ecosystem health. Given the vast number of untracked logging roads, a reliable method for detection and monitoring is essential. Manual tracing is both unrealistic and time-consuming, highlighting the need for an automated solution.

Artificial intelligence, particularly deep learning, has shown promise in tackling this problem. Previous research has demonstrated that segmentation models can be effectively used to detect logging roads through satellite imagery. In our work, we aimed to explore the viability of these methods for detecting logging roads in New Brunswick. Our project involved training various segmentation models on a small dataset to detect logging roads in North-Western New Brunswick.

This study serves as an exploratory effort to assess the feasibility of using deep learning for this purpose. We worked with a limited dataset of nine training images and four validation images, which were further cropped to create 900 training images and 500 validation images across four different crop sizes. Our training results indicated that the models struggled to converge, with

F1-scores often stagnating below 55%. However, upon reviewing the visual outputs, the models showed promising potential. While the models effectively detected the roads, they frequently generated excessive noise, which can be attributed to the limited quality and quantity of the training data.

Despite these challenges, our results indicate that this approach holds promise, and future work—particularly with larger, higher-quality datasets—could yield significant improvements in performance.

Proposed Future Work

The scope for future work on this project is vast, encompassing improvements in algorithms, techniques, processing methods, and datasets. A key priority is the enhancement of our training datasets. As noted earlier, the current dataset is relatively small, which significantly affects the quality of the training process. Expanding the dataset to include a few hundred to thousands of images would likely lead to substantial performance improvements. Moreover, increasing the quality of the datasets is equally crucial. Since the performance of our models is directly tied to the quality of the data, a subpar dataset will inevitably result in subpar models.In terms of models and algorithms, our current exploration has been limited to the U-Net architecture with 22 different algorithms. Future work should involve testing additional architectures, such as LinkNet and PAN, which could offer interesting alternatives for road detection tasks.

As a roadmap, we propose the following timeline to develop a robust solution for detecting roads in satellite images:

1. **Dataset Generation and Verification**: The first step is to create a high-quality dataset composed of hundreds of large images, ideally of size 5000x5000 pixels or larger. The dataset must not only be extensive but also meticulously verified for accuracy, ensuring that the models can learn effectively from reliable ground truth data. Accurate labeling is crucial for producing reliable predictions, and the dataset should cover a variety of environments to improve generalization. 1-2 months would be required.

2. **Comprehensive Study of Model Parameters**: Next, a thorough investigation of algorithms, model architectures, optimizers, cropping sizes, and augmentation techniques needs to be conducted. While our initial tests were carried out with a limited set of parameters, further research will involve experimenting with thousands of training iterations to identify the optimal combinations. This step would require a systematic approach, involving hyperparameter tuning and cross-validation to refine the models and boost performance. 2-4 months required.

3. **Advanced Noise-Reduction Techniques**: Finally, noise-reduction techniques will need to be explored in greater depth to fine-tune the predictions and improve overall accuracy. Post-processing methods, such as connected components, morphological operations, and more advanced approaches like Conditional Random Fields (CRFs) or attention-based mechanisms, should be examined. Fine-tuning these techniques will help reduce false positives and enhance the precision of road segmentation in noisy or complex regions. 1-2 months required.

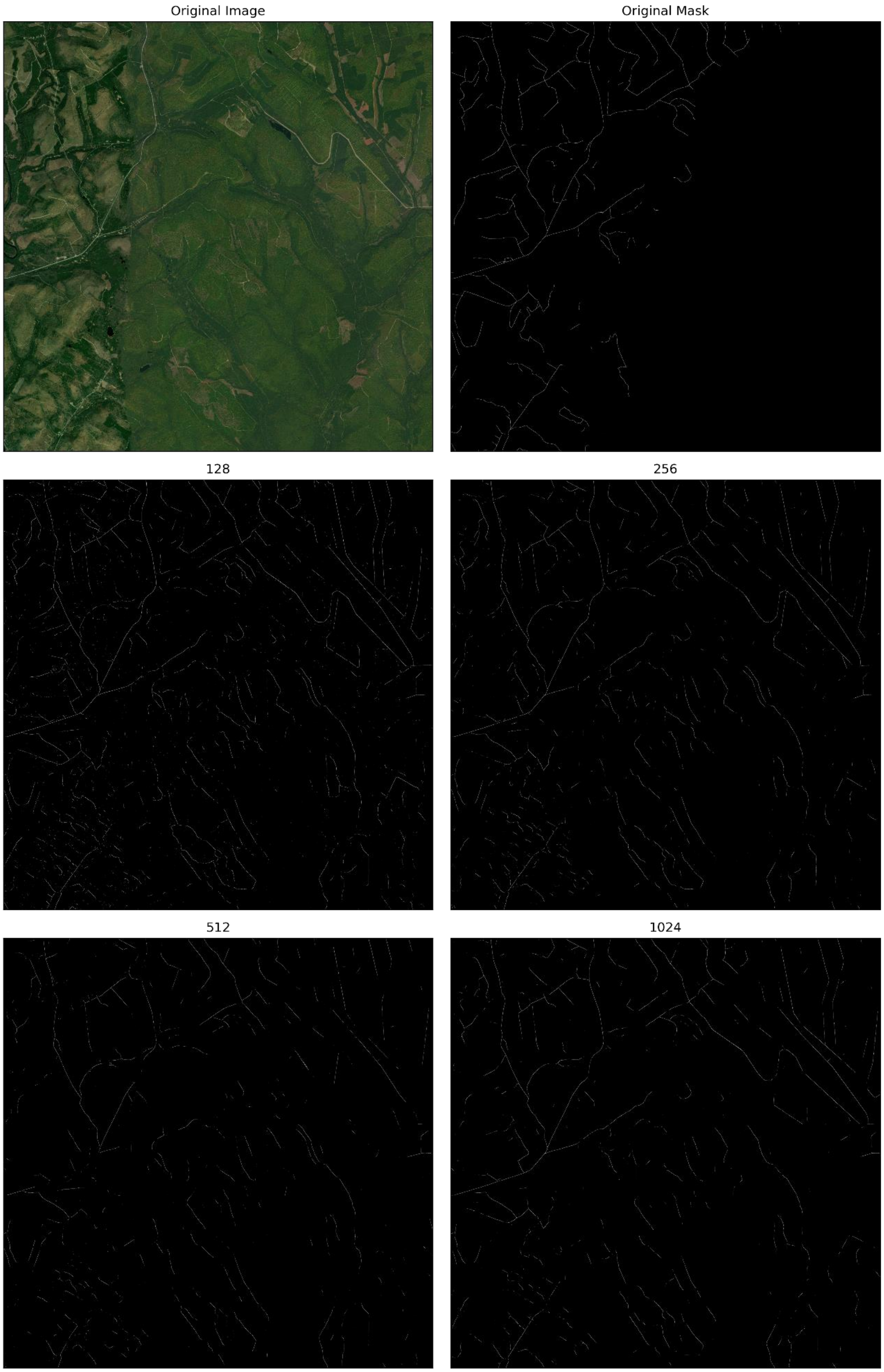

Figure 1: Full Image Comparison

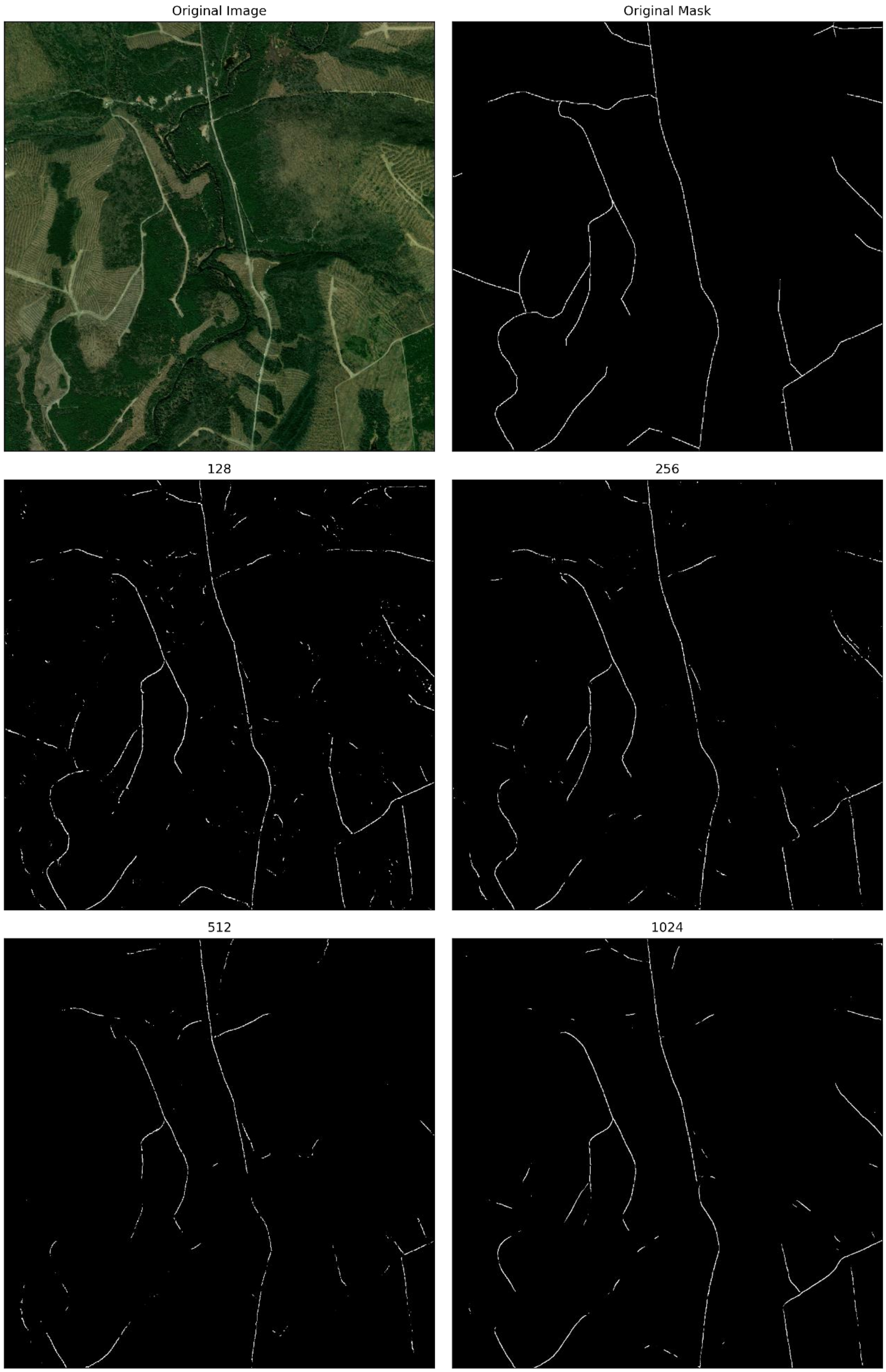

Figure 2 Cropped Image Comparison

Bibliography

Anderson, P. G. (1996). Sediment generation from forestry operations and associated effects on aquatic ecosystems. *Proceedings of the Forest-Fish Conference: Land Management Practices Affecting Aquatic Ecosystems, 5*.

Ayala, C. a. (2021). A deep learning approach to an enhanced building footprint and road detection in high-resolution satellite imagery. *Remote Sensing, 13*(16), 3135.

Botelho Jr, J. a. (2022). Mapping roads in the Brazilian Amazon with artificial intelligence and Sentinel-2. *Remote Sensing, 14*(15), 3625.

Jutras, S. (2022, April 4). *How Québec's abandoned logging roads are damaging lakes, rivers and streams.* Retrieved from smartwatermagazine.com: https://smartwatermagazine.com/blogs/sylvain-jutras/how-quebecs-abandoned-logging-roads-are-damaging-lakes-rivers-and-streams#:~:text=Built%20almost%20exclusively%20for%20the,Qu%C3%A9bec%20government%20in%20recent%20decades.

Kleinschroth, F. a. (2017). Impacts of logging roads on tropical forests. *Biotropica, 49*(5), 620-635.

McCashion, J. D. (1983). Erosion on logging roads in northwestern California: How much is avoidable? *Journal of Forestry, 81*(1), 23-26.

Mokhtarzade, M. a. (2007). Road detection from high-resolution satellite images using artificial neural networks. *International journal of applied earth observation and geoinformation, 9*(1), 32-40.

Natural Resources and Energy Development NB. (n.d.). *Crown Roads.* Retrieved from gnb.ca: https://www2.gnb.ca/content/gnb/en/departments/erd/forestry-conservation/content/crown-lands-forest-operations/crown-roads.html

Sloan, S. a. (2024). Mapping remote roads using artificial intelligence and satellite imagery. *Remote Sensing, 16*(5).

Our public github can be found here while our private github including all of our commits can be found here, please contact emb9357@umoncton.ca for access to the private repo.

Annex

We also trained several models using our V3 dataset, which we received in mid-September. Unfortunately, due to the late arrival of the dataset, we were only able to train a limited number of models. Nevertheless, we present preliminary results that incorporate noise-cleaning techniques such as erosion and dilation. Additionally, we trained our models with and without augmentation, and the results are visually demonstrated using DenseNet121 with a U-Net architecture.

First, we present the training results without augmentation for images 14 and 15 in Figures 3 and 4, respectively. Cropped versions of these images are shown in Figures 5 and 6. Figures 7 through 10 illustrate the results when augmentation techniques were applied. Since our predictions exhibited considerable noise, we propose introducing noise-reduction techniques specifically for the augmented predictions, as they consistently produced slightly better overall results.

Figures 11 to 18 display the predictions for image 15 at various sizes, comparing the results between full images and cropped sections. In general, larger cropped sections yielded better results than smaller ones, likely because the larger crops captured more spatial information, distinguishing roads from non-road areas more effectively. Moreover, larger cropped sections stitched together produced more continuous road segments, making them less susceptible to noise-canceling techniques.

Among the noise-reduction methods, connected components (CC) achieved the best results compared to contouring and morphological operations. However, these techniques were implemented quickly and have not been fine-tuned. Further refinement and exploration of additional noise-reduction techniques are necessary.

Lastly, we observed that the dataset, particularly the masks, still presents issues with quality. For example, some masks, like the one for image 15, are missing significant portions of roads, as shown in Figure 6. This lack of complete data negatively impacted model training, as the models were unable to learn from properly annotated ground truth.

Future Work and Improvements Additional

Looking forward, several avenues can be pursued to improve model performance and prediction quality:

1. **Advanced Noise-Reduction Techniques**: While connected components (CC) showed the best results in reducing noise, more sophisticated methods could further enhance prediction accuracy. Algorithms like Conditional Random Fields (CRFs) could be used for post-processing to better refine the boundaries of predicted road segments. Additionally, exploring techniques such as deep learning-based denoising or noise-removal autoencoders could provide further improvements.

2. **Fine-Tuning of Existing Methods**: The current noise-reduction techniques (contouring, morphological operations, CC) were applied without extensive fine-tuning. A more systematic approach to tuning these methods could yield better results, especially if optimized for different road types or environmental conditions (e.g., urban vs. rural roads).

3. **Augmentation Strategies**: While basic augmentation was applied, more advanced augmentation techniques could be explored. Approaches like mixup, cutout, or even using Generative Adversarial Networks (GANs) to generate synthetic training examples could diversify the dataset and improve model generalization.

4. **Better Dataset Quality**: The current dataset contains several issues, particularly in the mask annotations, which hinder model performance. A focused effort to clean and enhance the dataset, either by manually correcting errors or using semi-supervised methods to improve mask quality, would be highly beneficial. Furthermore, incorporating additional datasets with higher-quality annotations and more diverse landscapes could help the models generalize better. An issue seem to be caused by cities such as driveways, thus increasing the quantity of this data type could be beneficial.

5. **Incorporation of New Algorithms**: As mentioned in our earlier proposed improvements, exploring other state-of-the-art algorithms such as EfficientNet or ResNeXt combined with U-Net could lead to improved performance. In addition, attention-based mechanisms like Attention U-Net could be explored to focus the model on relevant features and reduce the impact of noise. Other architectures such as LinkNet and PAN could be interesting.

6. **Multi-Resolution or Multi-Scale Approaches**: To address the issue of spatial context, multi-resolution or multi-scale approaches could be applied. By training models to analyze images at different resolutions simultaneously, the network can capture both fine-grained details and broader context, potentially reducing noise in smaller crops while maintaining overall spatial accuracy. It could be beneficial for the models to have both crops of 128 and 1,024 into a single model to provide different scales, thus increasing the spatial information.

7. **Ensemble Learning**: An ensemble of models trained with different noise-reduction techniques, architectures, and data augmentations could further improve the robustness of the predictions. By combining the strengths of multiple models, we can achieve better overall accuracy and resilience to noise. Ensemble learning, or voting methods, have achieved good results in the past in this field.

8. **Additional Post-Processing Techniques**: Finally, in addition to noise-reduction methods, other post-processing techniques such as edge refinement, super-resolution techniques for road edges, and road topology-aware correction algorithms could be explored. These methods would refine the segmentation, particularly in challenging areas such as intersections or narrow road segments.

By following these future directions, we believe that the model’s robustness and accuracy can be significantly improved, leading to more reliable road segmentation, even in the presence of noisy or incomplete data.

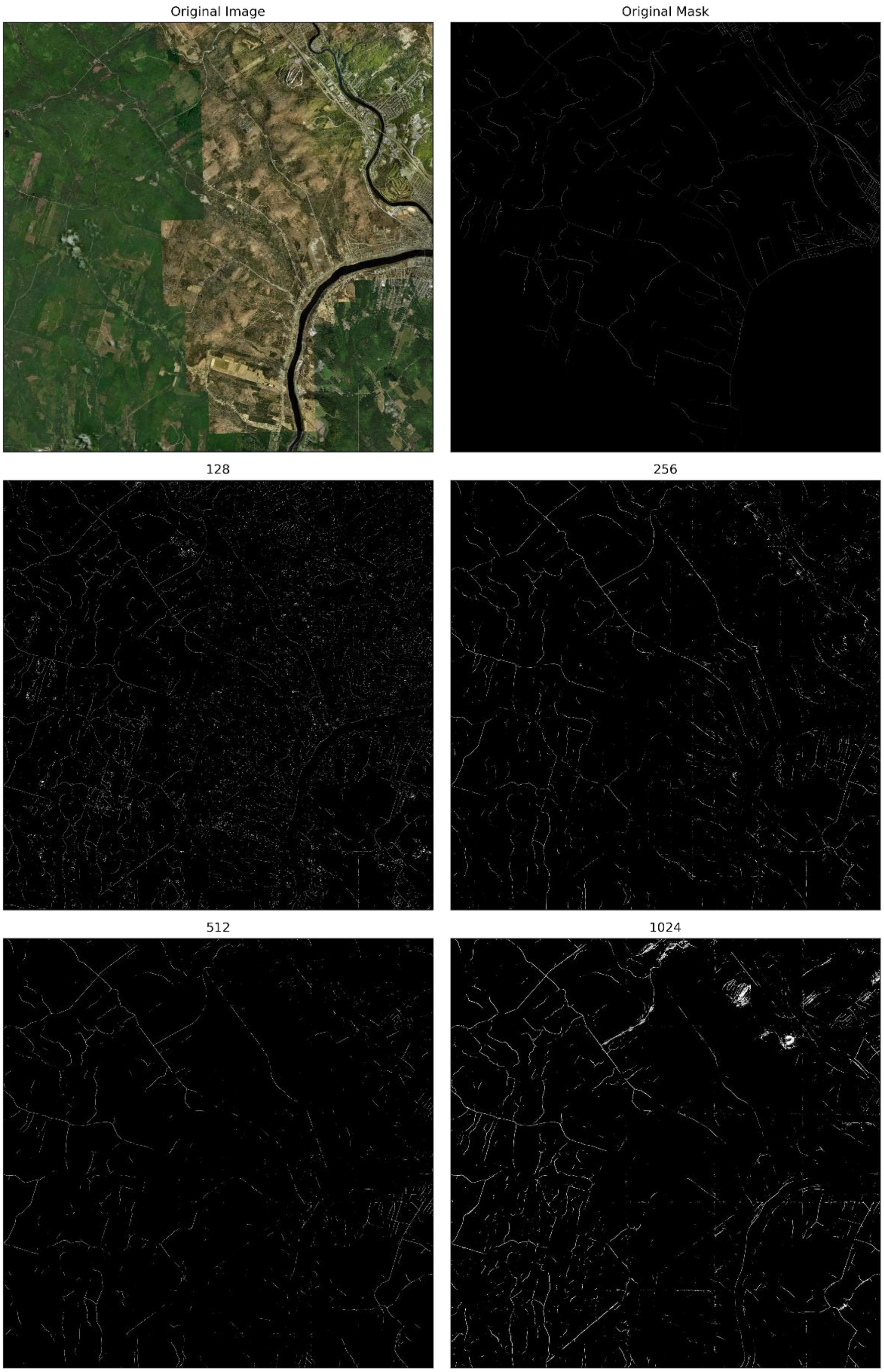

Figure 3 Image 14 No Augmentation

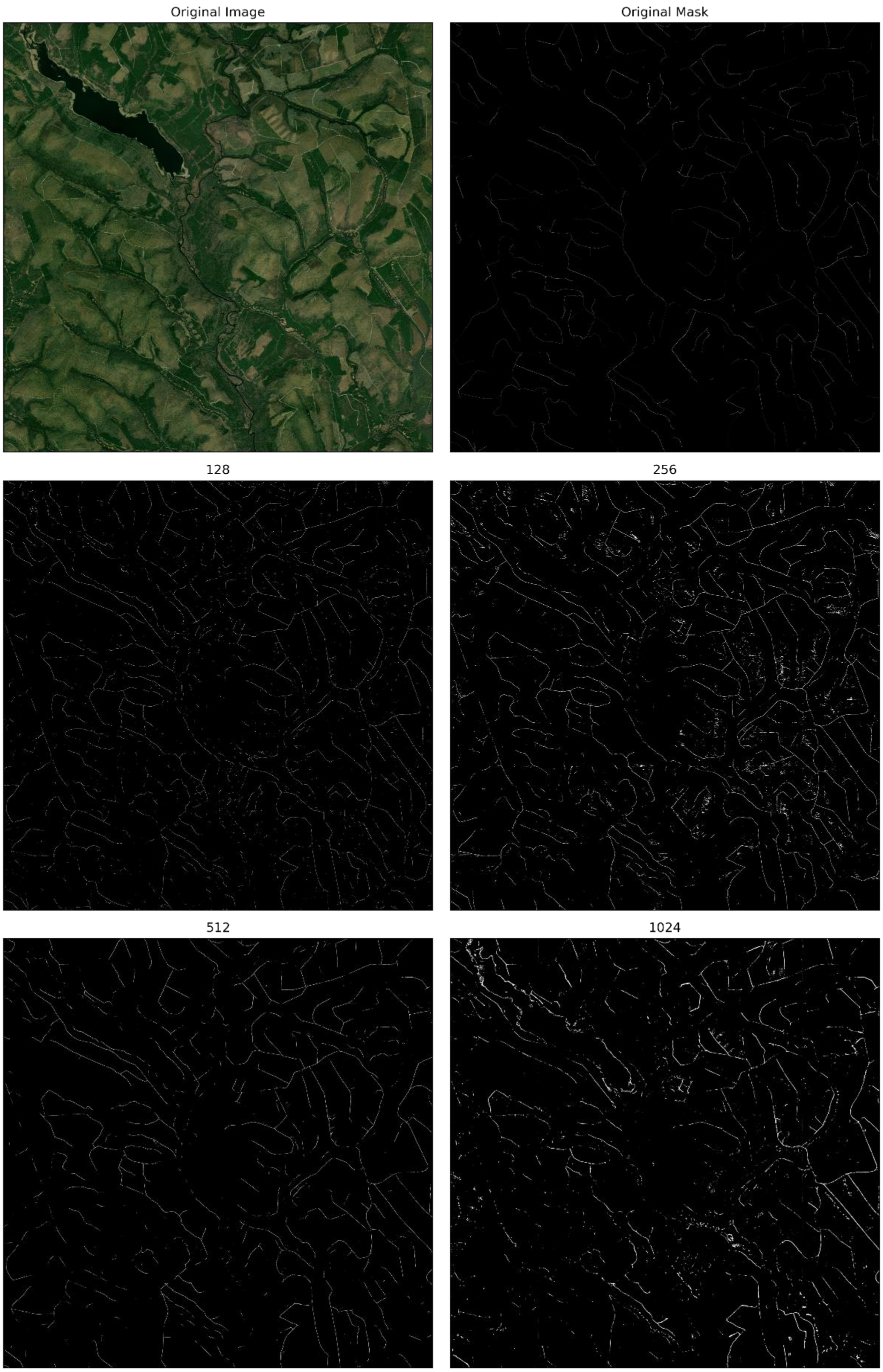

Figure 4 Image 15 No Augmentation

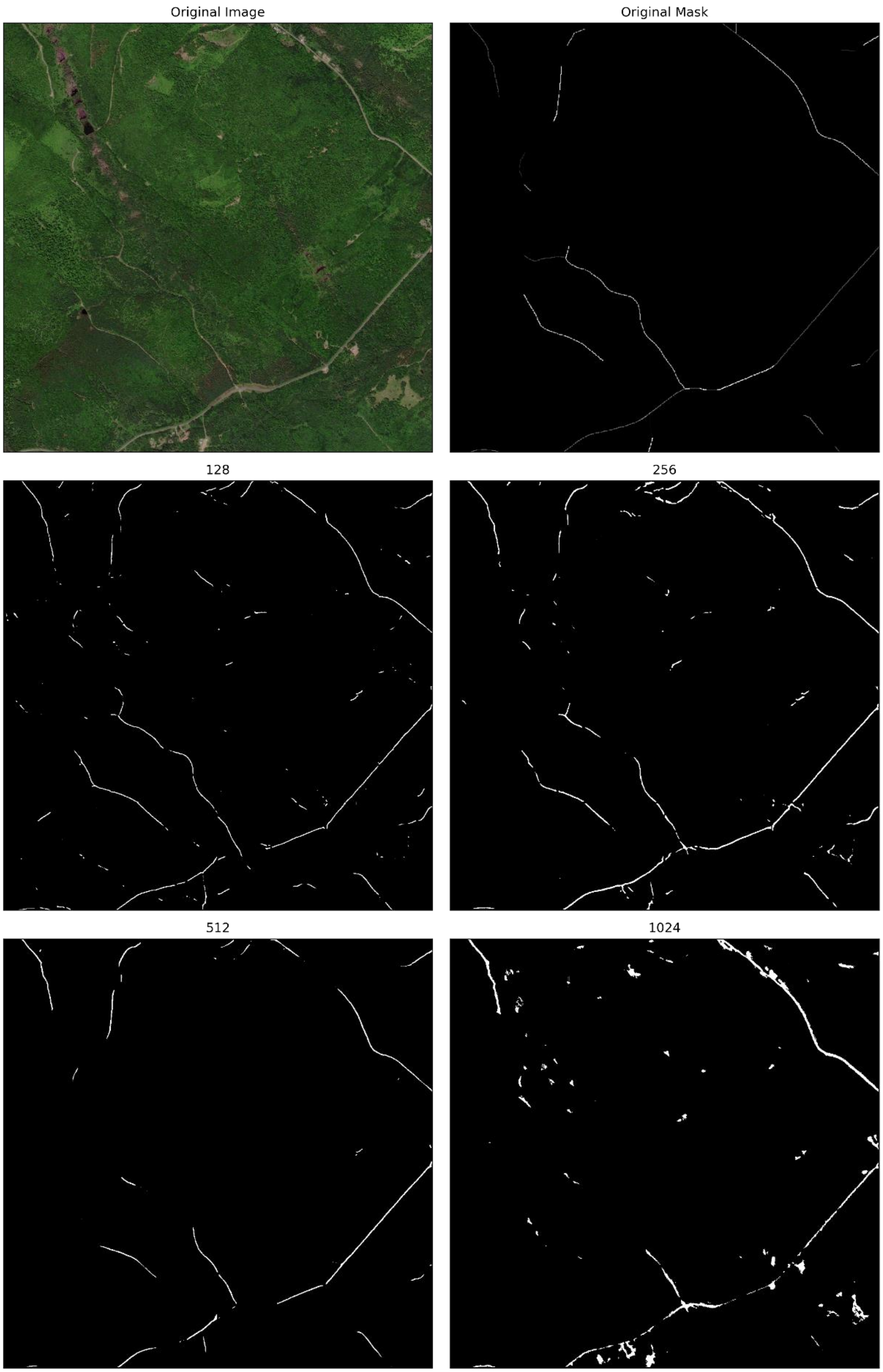

Figure 5 Image 14 No Augmentation Cropped

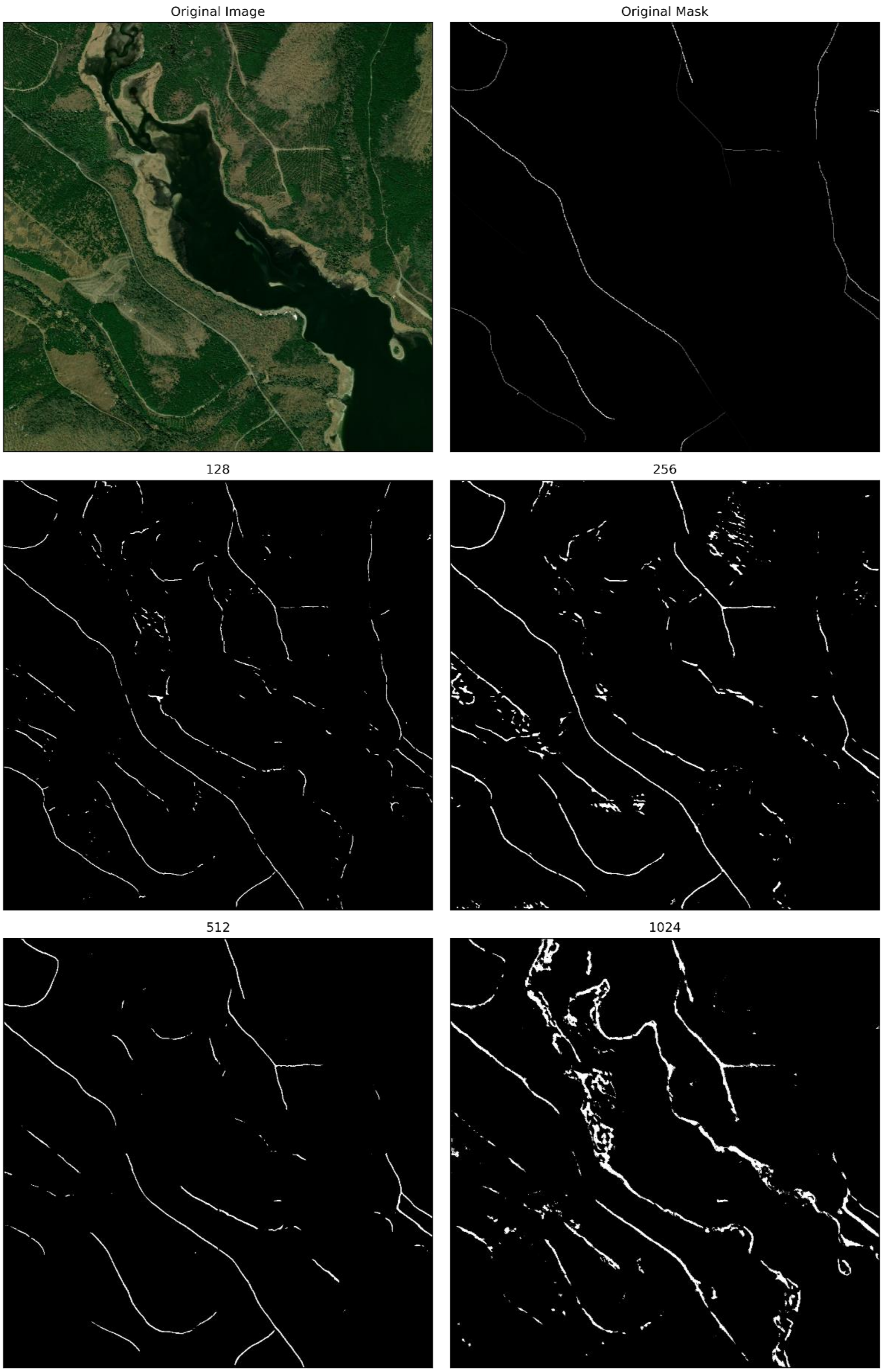

Figure 6 Image 15 No Augmentation Cropped

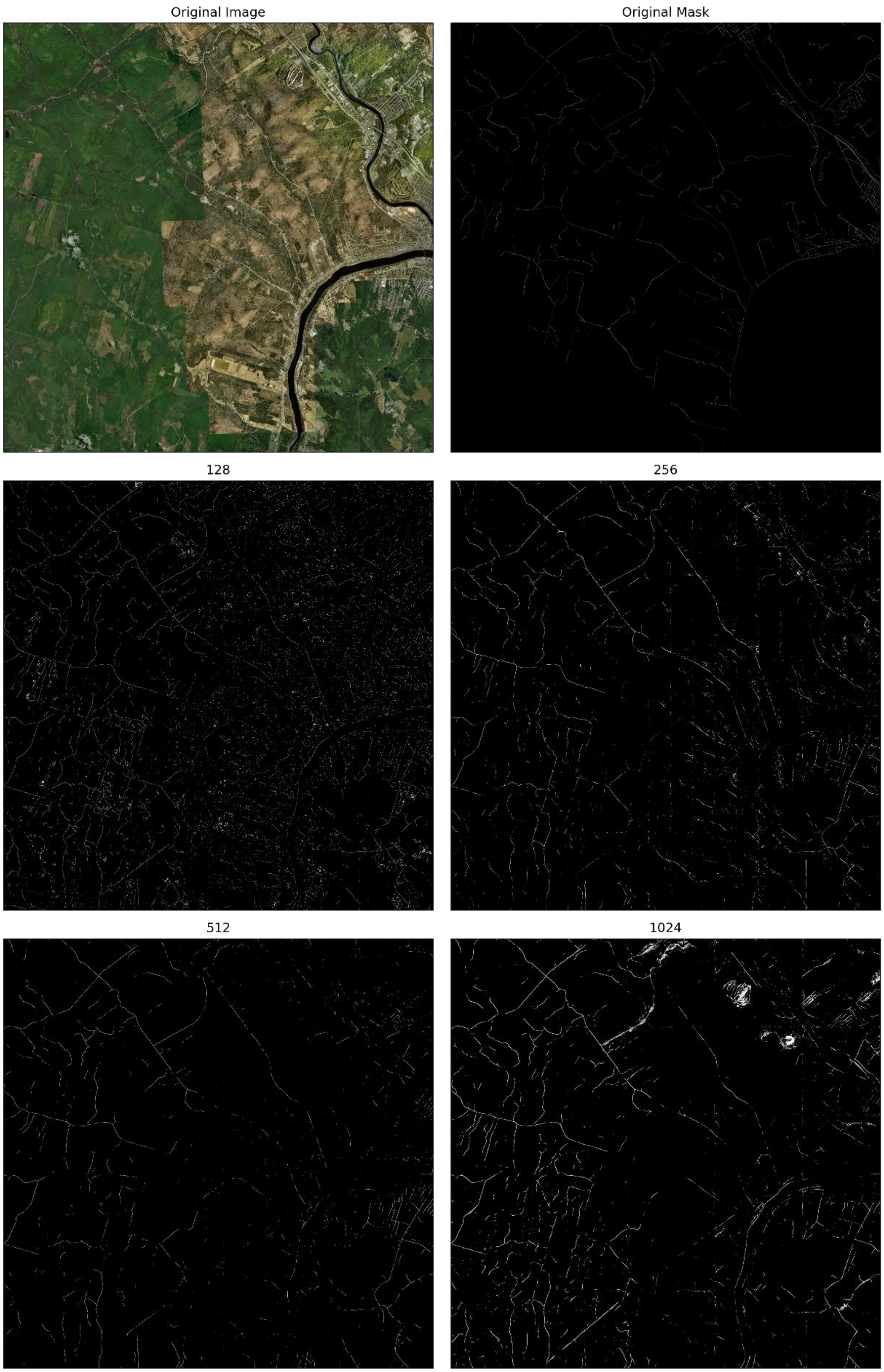

Figure 7 Image 14 Augmentation

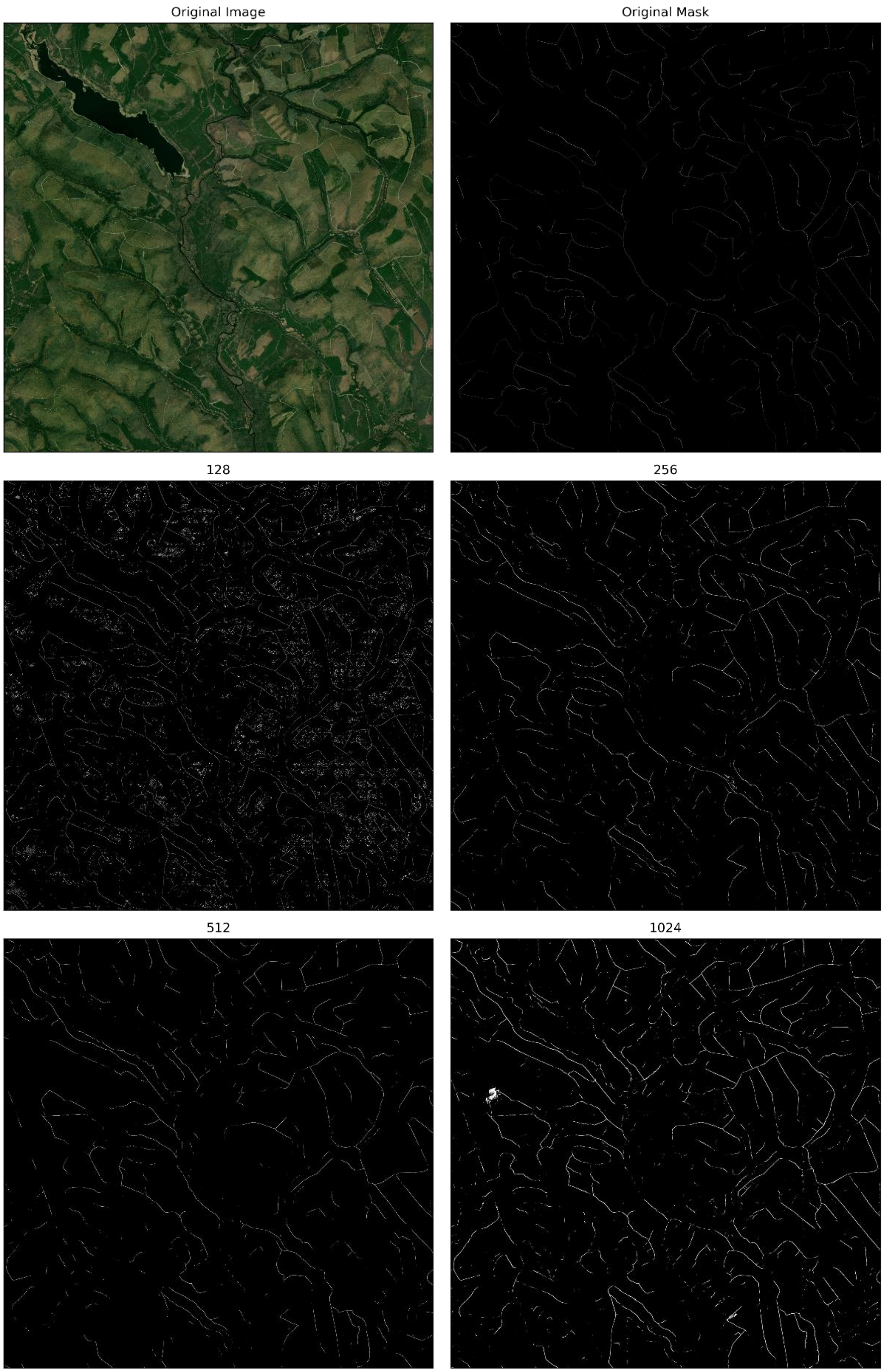

Figure 8 Image 15 Augmentation

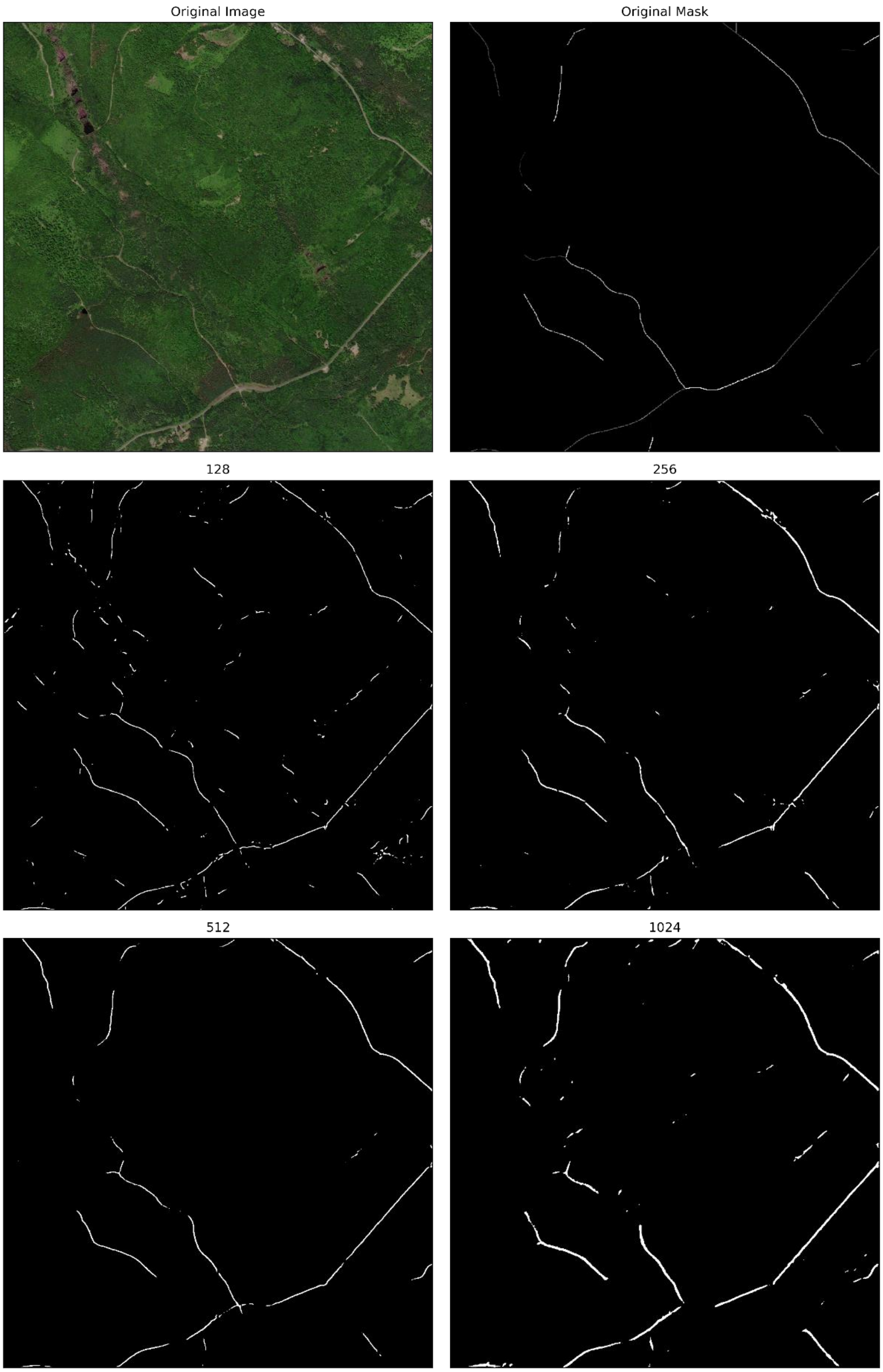

Figure 9 Image 14 Augmentation Cropped

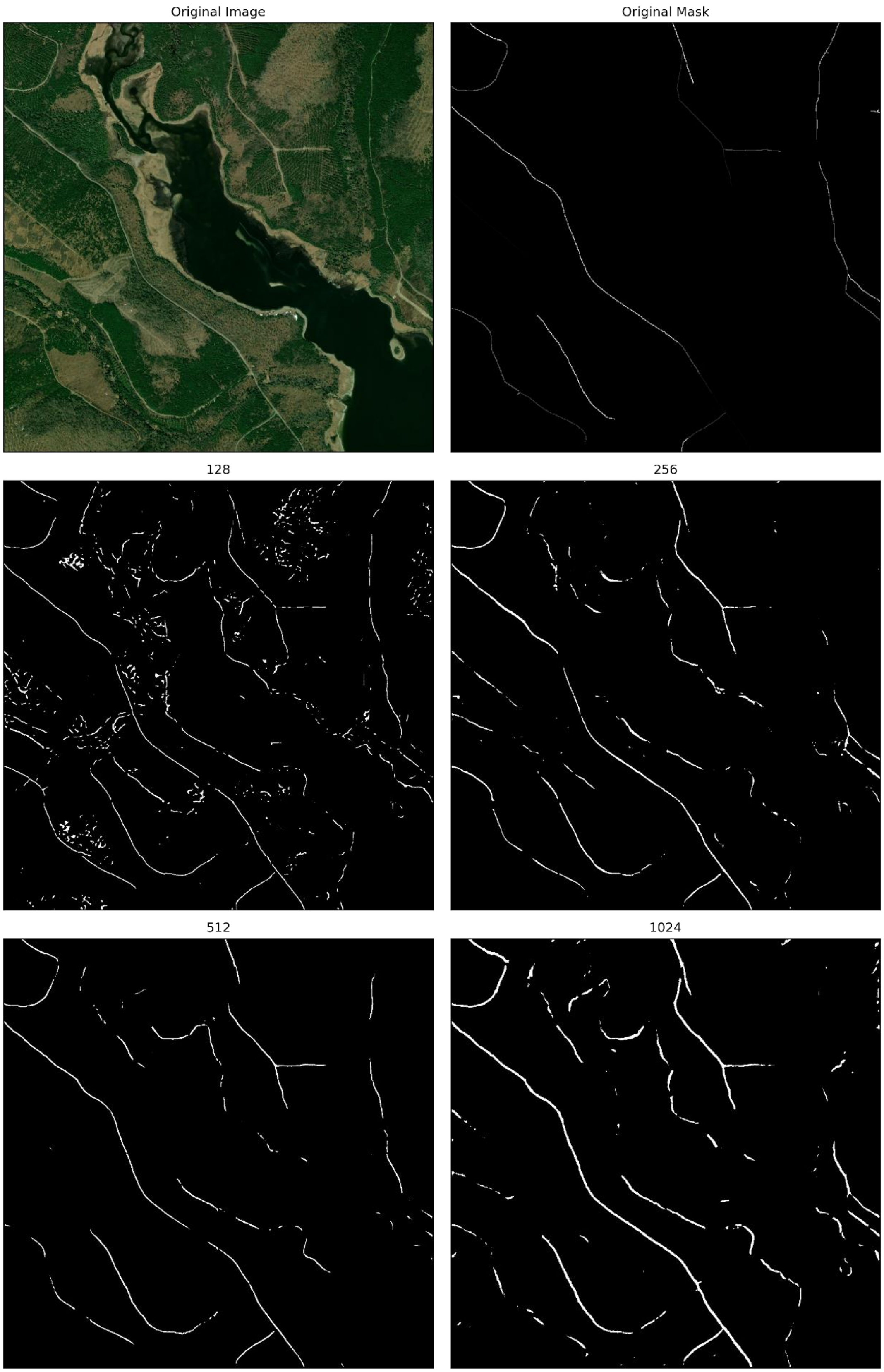

Figure 10 Image 15Augmentation Cropped

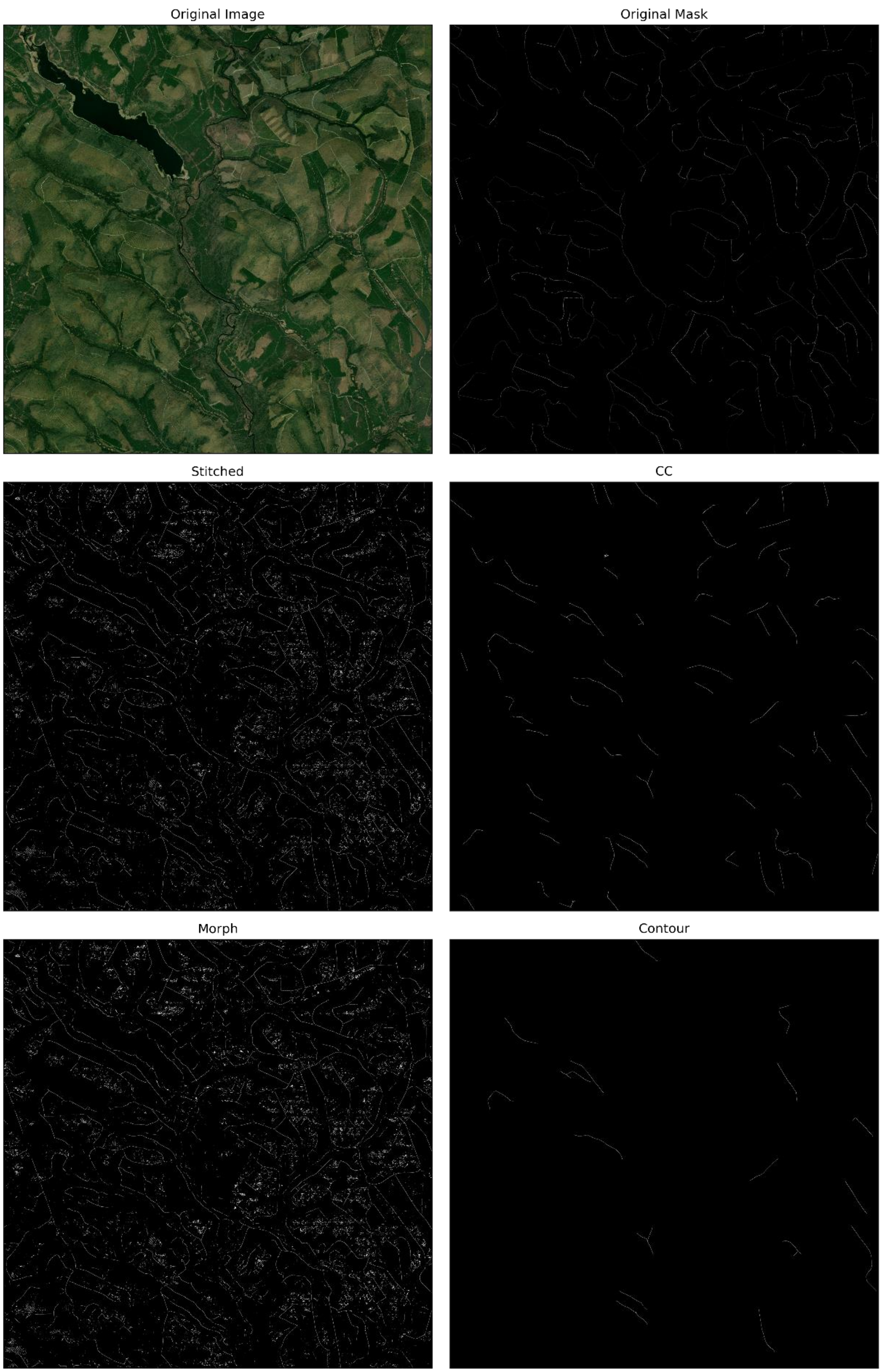

Figure 11 Image 15, size 128

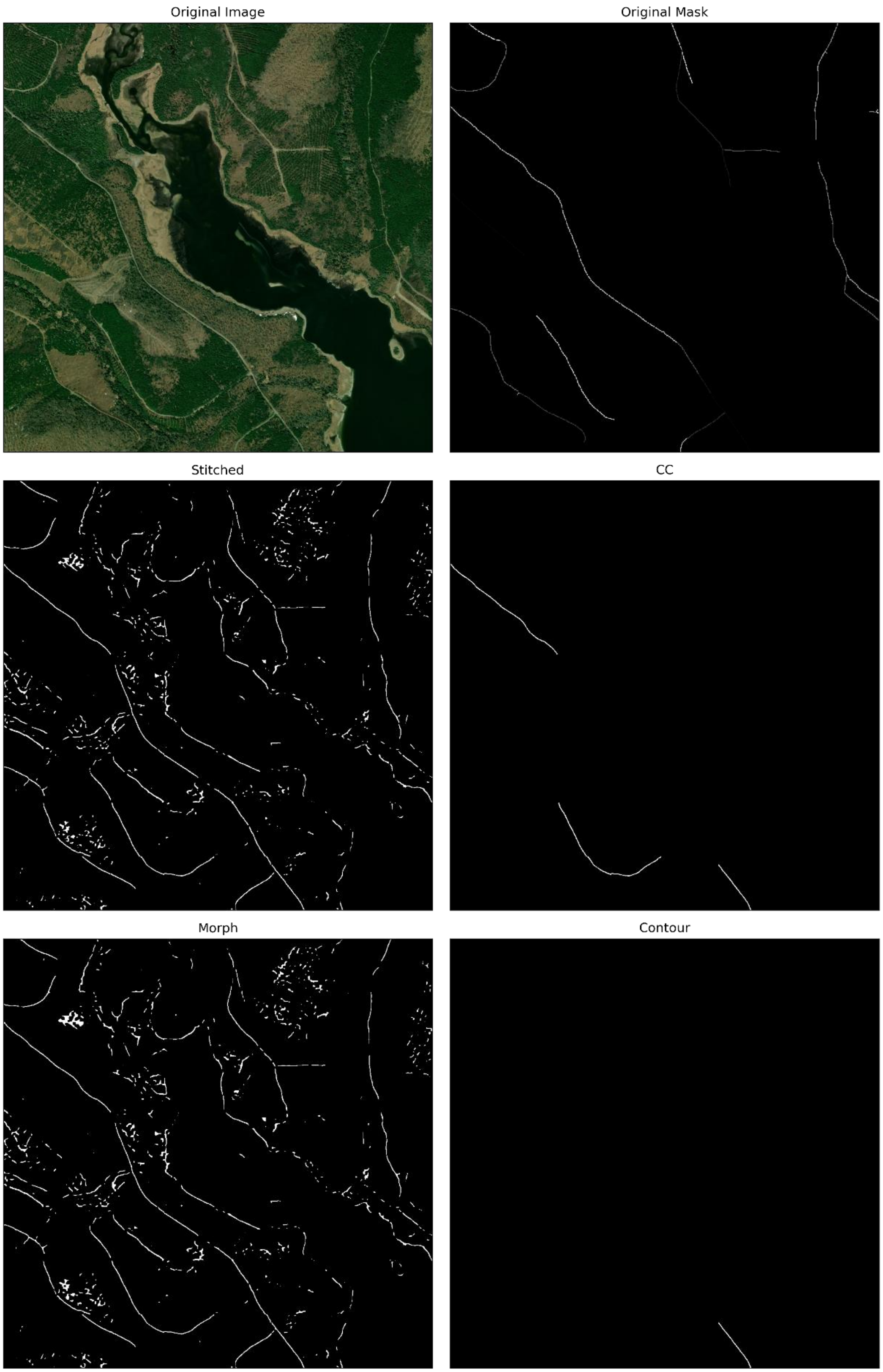

Figure 12 Image 15, size 128, Cropped

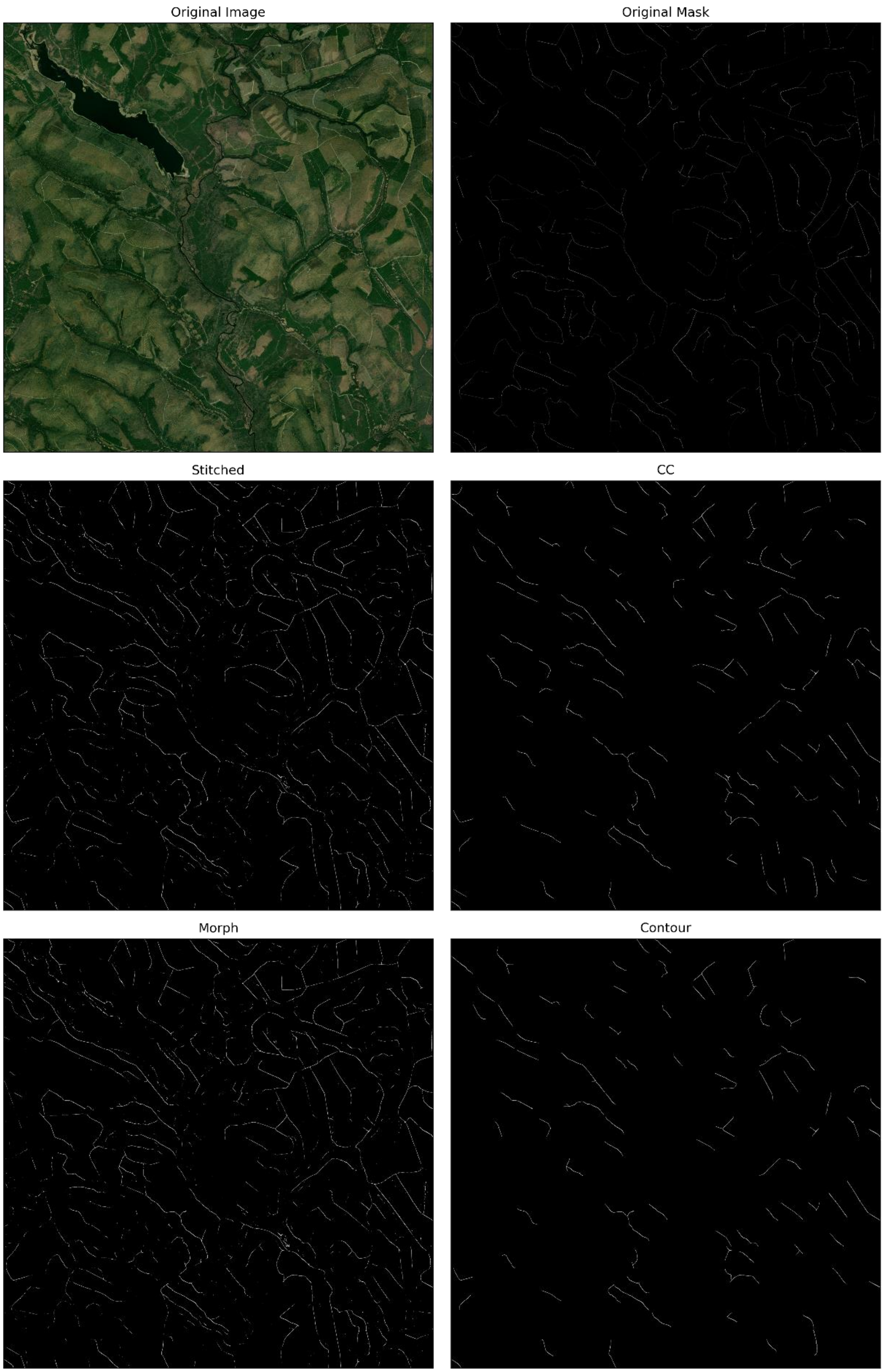

Figure 13 Image 15, size 256

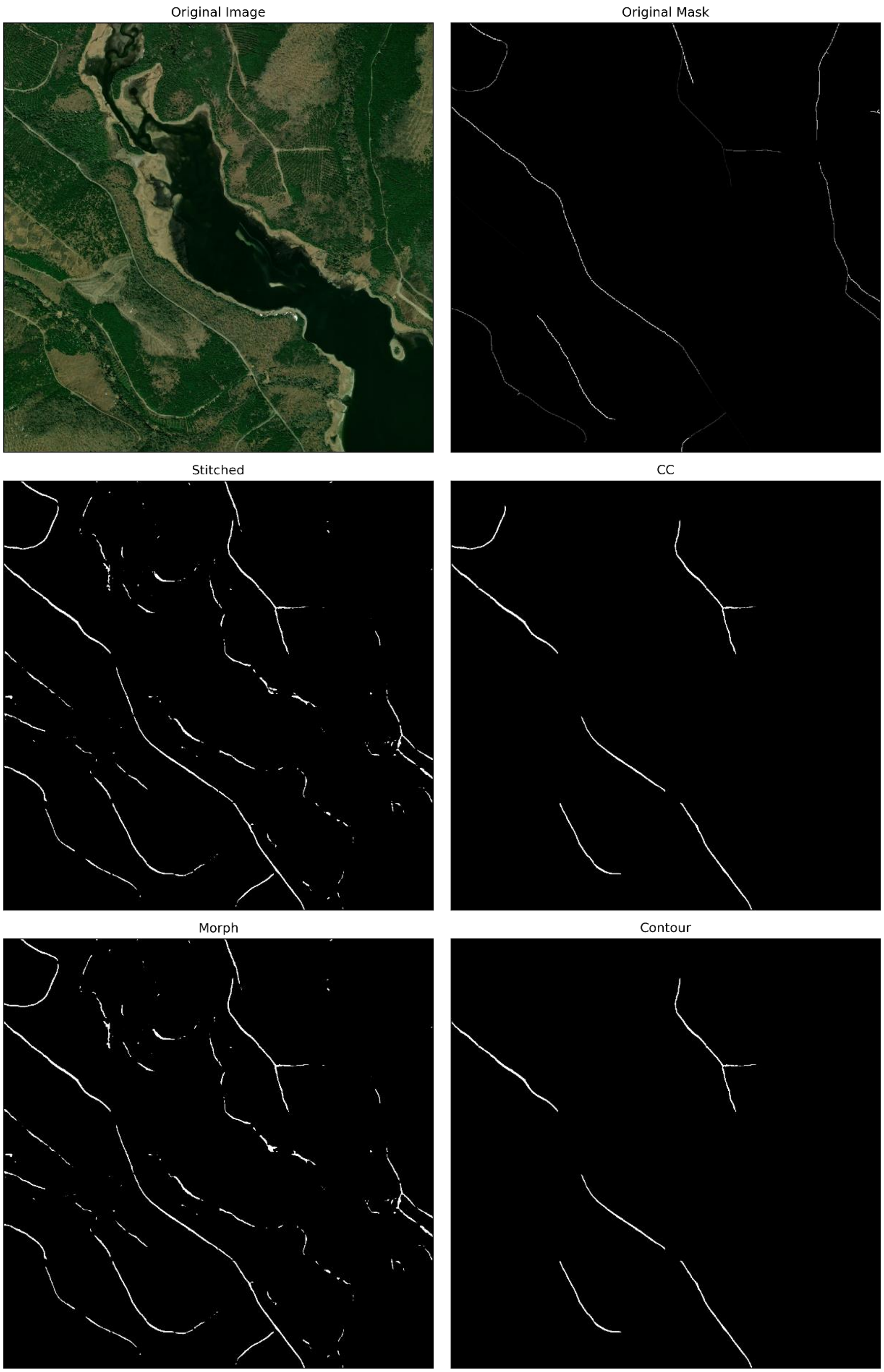

Figure 14 Image 15, size 256, Cropped

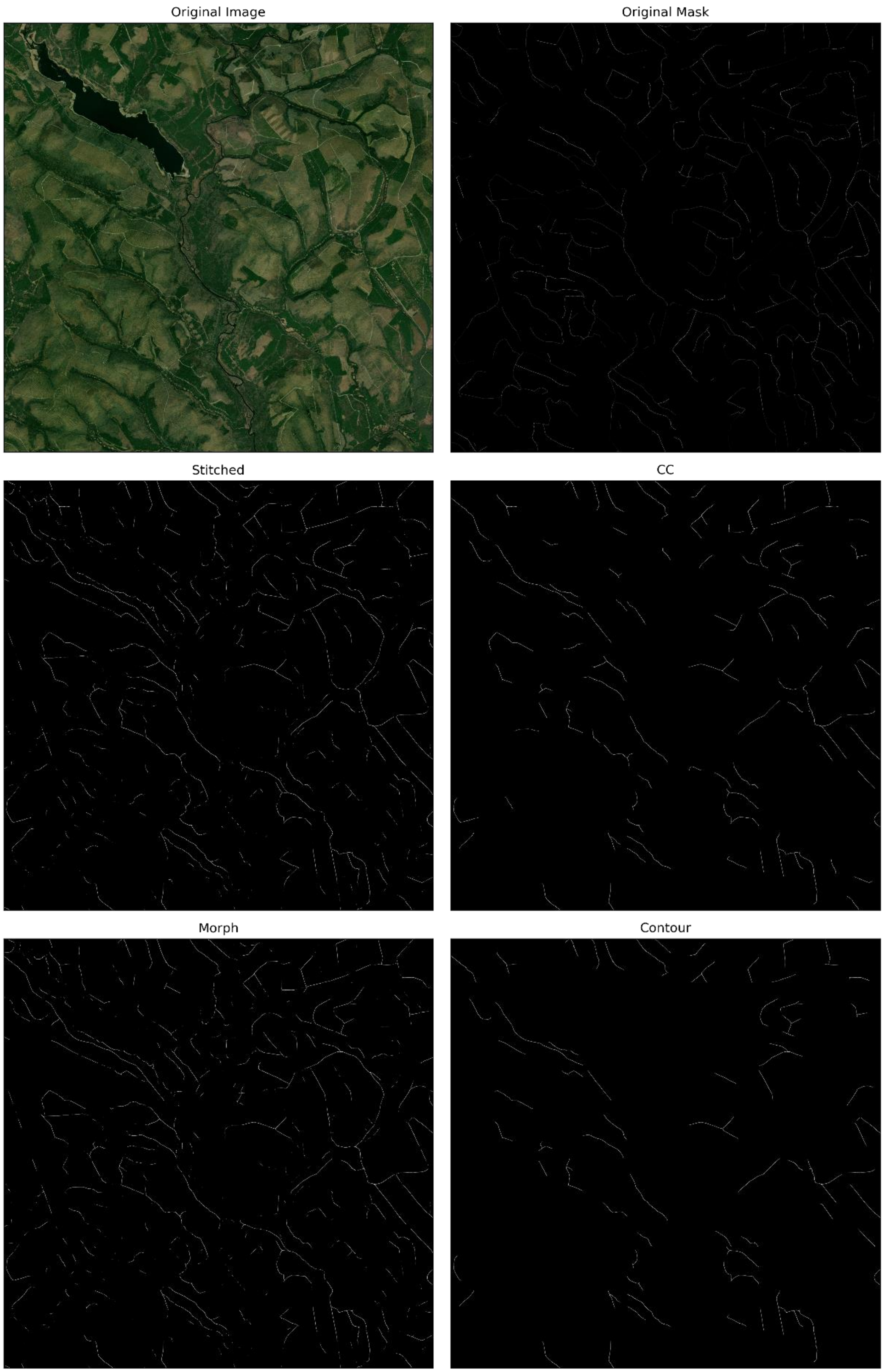

Figure 15 Image 15, size 512

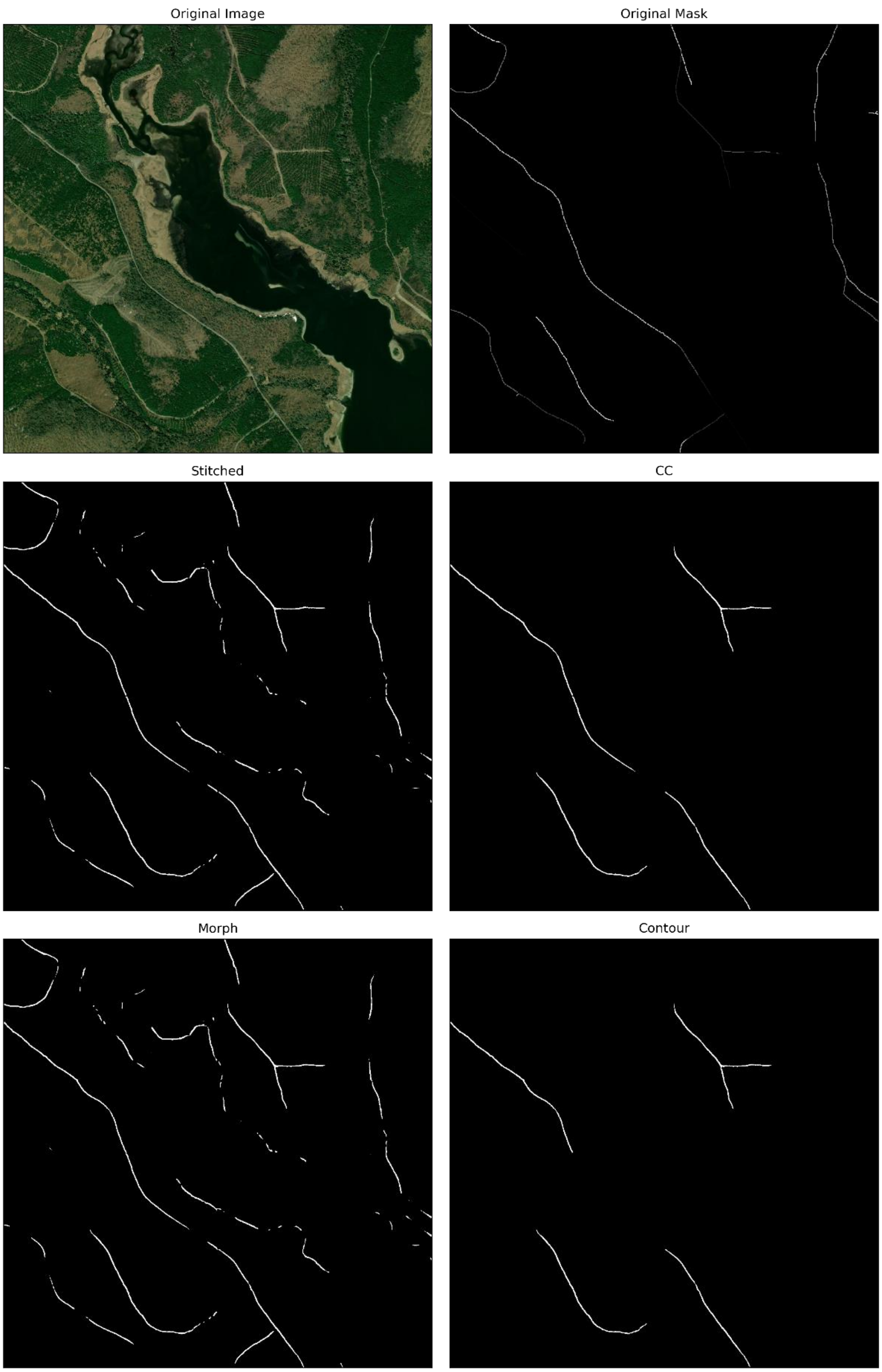

Figure 16 Image 15, size 512, Cropped

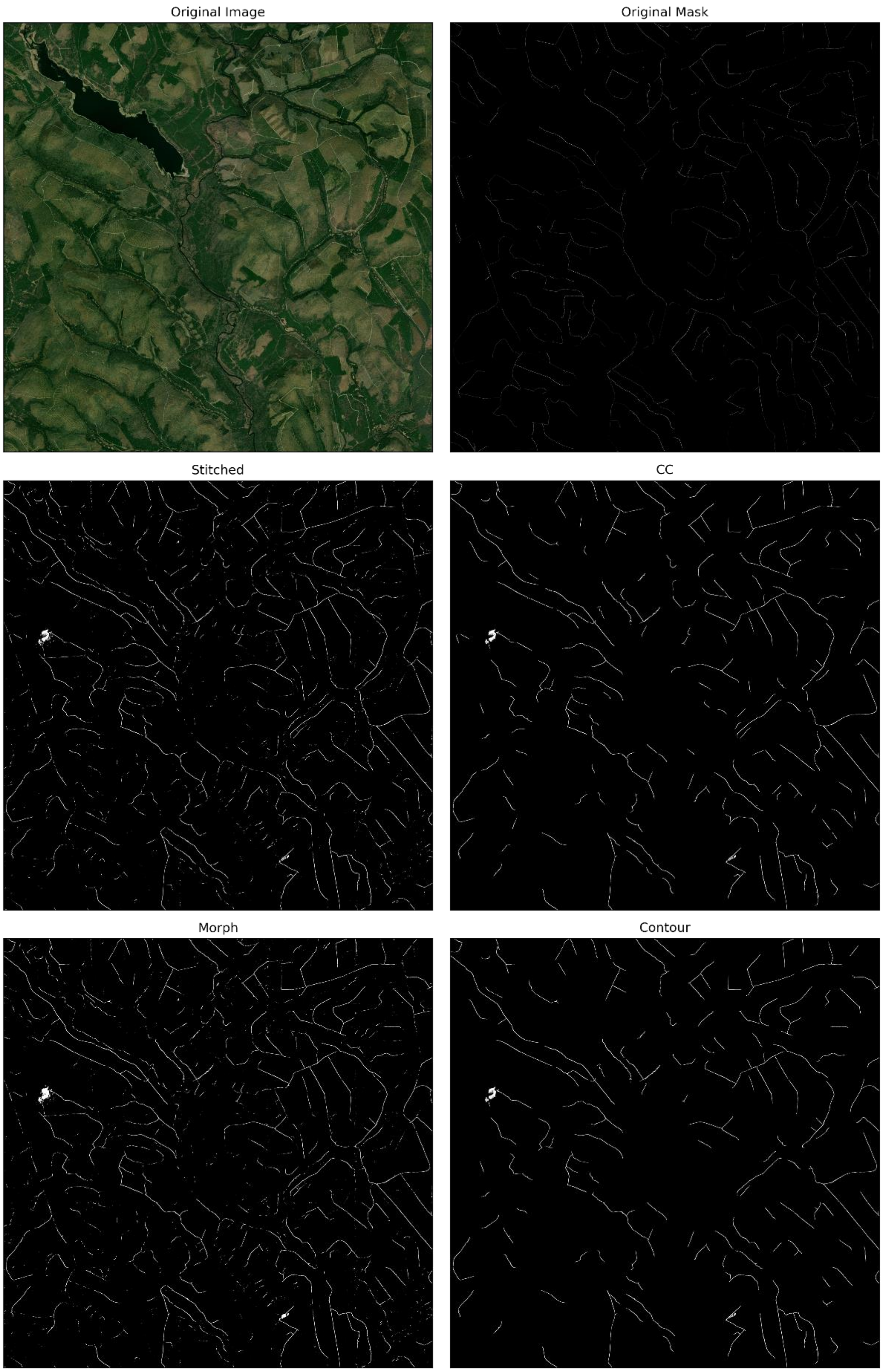

Figure 17 Image 15, size 1024

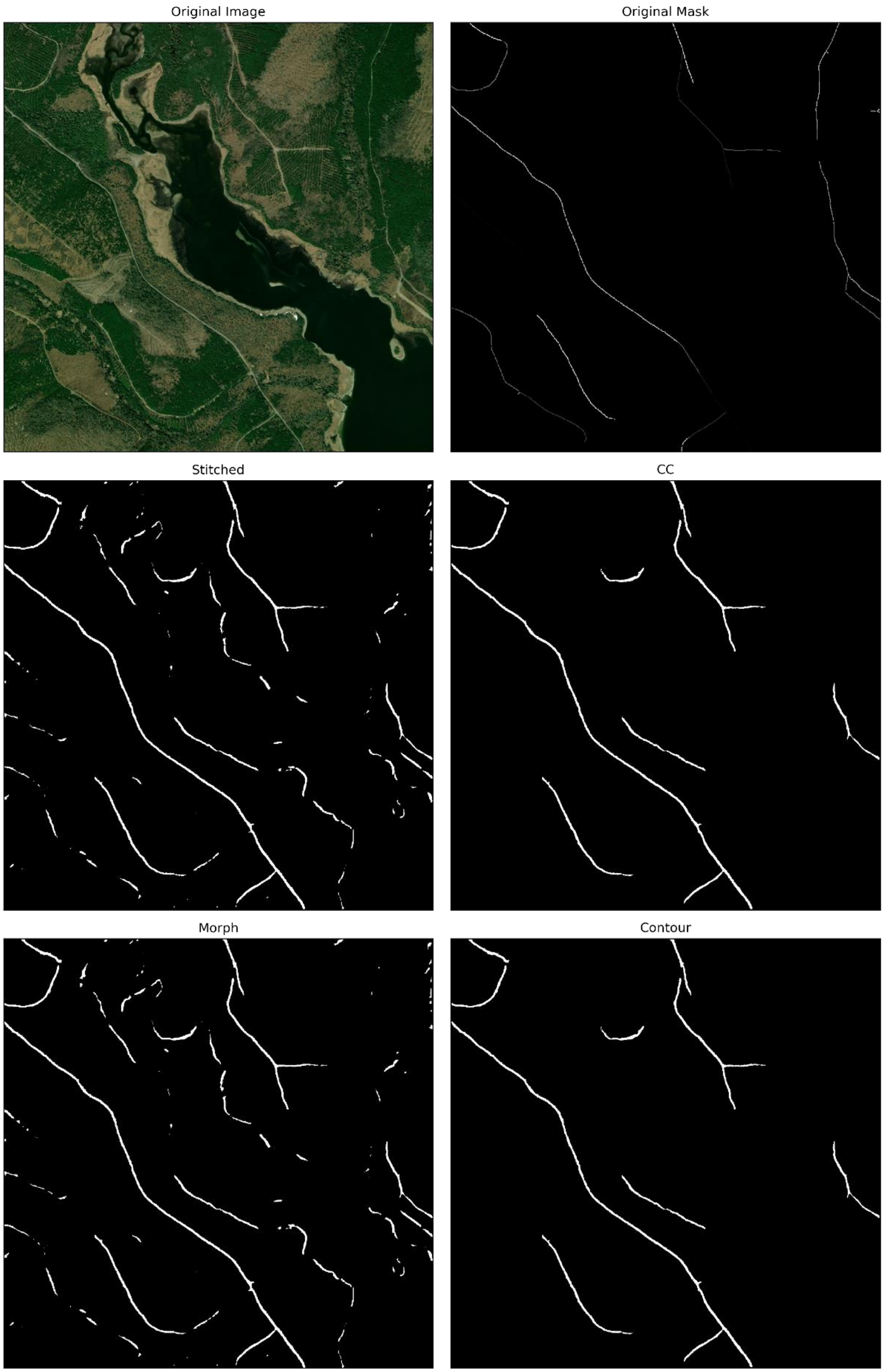

Figure 18 Image 15, size 1024, Cropped