

# Bare Demo of IEEEtran.cls for IEEE Journals

Michael Shell, *Member, IEEE*, John Doe, *Fellow, OSA*, and Jane Doe, *Life Fellow, IEEE*

**Abstract**—The abstract goes here.

**Index Terms**—IEEE, IEEEtran, journal, LATEX, paper, template.

## I. INTRODUCTION

### M: Motivate need for powerline inspection

Electricity is the lifeblood of modern society, thus, it is essential to ensure stable electrical power supply across the nations. Hence, it is of utmost importance for utilities to inspect and maintain their electrical facilities on a regular basis. These tasks were conventionally done by human inspector manually following, observing and assessing the power grid. Recently, Unmanned Aerial Vehicles (UAVs) have been deployed to produce more quality and efficient observations due to their superior speed and ability to access higher altitude and more hazardous environments. To further boost the effectiveness and efficiency of inspection and maintenance tasks, the assessment step can be done automatically from the observations of UAVs, which, in many cases, are RGB images from cameras. Many automatic assessment solutions have recently been proposed, many of which are fueled by the power of artificial intelligence and deep learning. Examples are (insulators, transmission towers, defects, severity). Among these tasks, cable, wire or powerline recognition and localization is very crucial. Accurate detection of powerlines helps focus the scope of downstream solutions for diagnosing faults (Tears, kinking, bird caging, etc...), identify intrusions (vegetation encroachment, etc.) on said powerlines. These tasks are extremely vital as failure in catching these undesirable events can lead to catastrophic and fatal consequences [?]. Furthermore, the ability to detect powerlines is imperative for UAVs, and other forms of low-altitude flights, to navigate safely.

However, it is nontrivial to detect powerlines due to their inconspicuous appearances. The powerline can be very thin and might be missed by proximity sensors on UAVs. On camera images, the width of powerlines can be as thin as 1 pixel. Furthermore, cluttered backgrounds and hindered powerline visibility and same-colored background, such as snow and white coating, can cause the powerlines to be imperceptible. Other conditions, such as fog and lighting, can also negatively affect the discernability of powerlines on images.

### M: Current solutions leverage DL

M. Shell was with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332 USA e-mail: (see <http://www.michaelshell.org/contact.html>).

J. Doe and J. Doe are with Anonymous University.

Manuscript received April 19, 2005; revised August 26, 2015.

**M: Current state-of-the-art:** Motivate shortcomings.  
Overlapping lines, can include a picture of it to motivate it further.

**M: In this work, ... (we propose the solution).**

**T**HIS demo file is intended to serve as a “starter file” for IEEE journal papers produced under LATEX using IEEEtran.cls version 1.8b and later. I wish you the best of success.

mds  
August 26, 2015

### A. Subsection Heading Here

Subsection text here.

1) *Subsubsection Heading Here*: Subsubsection text here.

## II. RELATED WORK

The problem of powerline detection has been tackled with traditional computer vision. An notable early attempt was proposed by Kasturi et al. [1], where Steger’s method [2] was used to extract features, which are then filtered using Hough transform to exclude short lines. Yan et al. [3] proposes Radon transform to generate line segments, group these segments by slope and distance thresholder, then finally use Kalman filter as the post-processing step. Li et al. [4] removes clutters and noises in the background using a pulse coupled neural filter before using Hough transforms and K-means clustering to detect and combine line segments. The aforementioned solutions and some others [5], [6], [7], [8] hold the strict assumptions that powerlines appear straight and in parallel orientations, and thus, do not always apply in reality. Song et al. [9] was able to cover curve powerlines by using a normalized graph cut model to link line segments, which were produced from the responses of a match filter and first-order derivative of Gaussian. However, the performance of all the methods above are affected by the conditions that the input images were taken in. Any differences in camera settings, environment, lighting, and view angle can lead to the hyper-parameters of these methods to be re-defined, which in turn may requires specialist expertise.

Overall, deep learning can help alleviate the problems that traditional computer vision face. With enough data, deep learning solutions for classification and detection can generalize well to many practical scenarios. Pan et al. [10] suggests training a convolutional neural network (CNN) that takes in edge features, which are produced by steerable filters, and classifies whether square patches from input image contains a line or not. Then Hough transform is used to detect the

powerline segments. Similarly, Gubbi et al. [11] also proposes using a CNN but, instead, uses Histogram of Gradient features as input. The line segment detector [12] was used as the post processing step. Since the inputs of these two methods are individual patches, the CNN models lack the contextual information of the entire images. Hence the performance can be limited, especially for low-contrast images with cluttered background. More recent approaches [13], [14], [15], [16], [17], [18] frame powerline detection as a pixel binary classification problems where each pixel is classified whether it belongs to the powerline other not. This type of problem can be undertaken by deep learning semantic segmentation networks. Madaan et al. [13] different dilated convolutional neural networks are experimented for segmenting powerlines. Yetgin et al. [16] finetuned a CNN, which was pretrained on ImageNet [19], with a newly initialized softmax layer that classifies the powerline existence. After that, features from immediate layers of the finetuned CNN are used to train another classifier to perform powerline segmentation. In [15], a CNN is introduced with two components: information fusion module and attention module. Information fusion module is in encoder-decoder structure, where decoding stages are combined with their corresponding same-scaled encoding stages to fused semantic and location information for accurate powerline segmentation. Abdelfattah et al. [17] trained a generative adversarial network (GAN) to generate modified version of input images where the powerline are highlighted. A semantic decoder connecting to an immediate layer of the generator is also trained jointly in order to perform the actual segmentation task. These semantic segmentation methods have achieved satisfying results, however, they require training data with pixel-level annotation, which can be laborious to obtain. Especially in the case of powerlines, which can often be quite slender yet span across images, more meticulousness is required in the annotating process. Choi et al. [20] tries to mitigate this burden by approximating pseudo segment maps, which are produced using Visualbackprop algorithm [21] from a classification CNN. This classification CNN is trained only with patch labels. Lee et al. [22] trained a CNN on images with only image-based label. During inference, this CNN is used on patches of high-resolution images and produces features map which can be combined for powerline localization.

**M: Provide high-level overview of current powerline inspection approaches and LS-Net. Mention shortcomings**

### III. METHODOLOGY

#### A. Preliminaries: LS-Net

LS-Net [23] was proposed as a single-shot line-segment detector inspired by SSD [?] and YOLO [?].

LS-Net reduces the need for pixel-level annotated data ground truth, which can be laborious to obtain. Especially in the case of powerlines, which can often be quite slender yet span across images, more meticulousness is required in the annotating process. Instead, LS-Net relies only on polyline annotation that trace along the powerline. Polyline annotation is much less effort and arguably more robust to labelling inaccuracy allowing more ground truth to be acquired facilitate better deep learning models.

LS-Net breaks down the problem of powerline detection into detecting and locating line segments in four overlapping grids, which are superimposed onto the input image. In each cell of the grids, LS-Net detect whether there exists parts of powerlines within its borders and provides the coordinates of the line segment endpoints. These divided results can be combined to produce a complete segment map of powerlines. The four-grid approach was proposed as opposed to one-grid approach, such as SSD [?] and YOLO [?], in order to encourage more thorough detection and localization of powerlines and combat the problem of discontinuities at grid cell borders and corners. This is illustrated in ??.

To perform line segment detection, the architecture of LS-Net involves a fully convolutional feature extractor which branches out into a classifier module and a regressor module as shown in ?? . The modules is to detect line segment existence and line segment endpoint coordinates respectively. The feature extractor of the original LS-Net was proposed to be a version of VGG-16 network [?] which was modified to include Group Normalization [?] before activations and replace max pooling layer with stride 2 for convolutional layer. The classifier module uses the feature map from the extractor and follows up with a 2 x 2 convolutional layer with stride 1 resulting in four sets of 512-channelled feature maps, in which each element in the feature map correspond to a cell of the respective grid. Finally, a 1 x 1 convolutional layer reduces the feature maps to a 2-channelled output. The channels correspond to the classes "line segment exists" and "line segment does not exist". The regressor module follows almost the identical design but ultimately produces a 4-channelled output instead, corresponding to the xy-coordinates of the predicted line segment endpoints. A detailed configuration of LS-Net is shown in ??.

Loss ???

#### B. LS-Netv2

**M: Instead of mentioning what we propose, we need to motivate it. Start from the problem (overlapping power lines) and use it to motivate the solution. Have a figure illustrating the whole pipeline and briefly describe the high-level method.**

**M: Then dive into the individual components (subsections), again starting from a motivation perspective**

It has been shown that LS-Net has been successful for cases included in the dataset [], which consists mostly images of powerlines being captured from a relatively close distance and the visibility of the powerlines throughout the span of the image. These cases might resembles distance of footage from autonomous line following drones flying in proximity with the powerline. However, in reality, images such as those used for visual inspection can be taken from further away making the some segments of powerlines appear really thin and/or blended with the background. If we follow the grid-based approach of LS-Net, the detection performance of powerline in each cell may depend on the information of areas around it. However, according to this calculation the theoretic receptive field of LS-Net is only four times the cell size (?) as calculated by

torchscan [?]), which LS-Net is unable to aggregate global context for the cell inference. Furthermore, the design of the first LS-Net is based on the assumption that only one power line segment is present for each division cell and thus only one segment is to be detected. This does not apply in practice where there exist powerlines appearing in close proximity and appearing to intersect each other (Figure ??).

To alleviate these limitations of LS-Net, we propose a new version of LS-Net, called LS-Netv2. In order to tackle the thin powerline problem, the first change we made is to increase the input size from 512 to 1536 (3x). This is to cater to the visual inspection pipeline of powerlines which has to often deal with high resolution images from far away. We discovered that resizing these high resolution images to 1536 is the sweet spot where the trade-off between powerline visibility and model efficiency. Using the same compression ratio, the size of each grid cell becomes 96.

As mentioned, the VGG-inspired backbone design results in a receptive field of about four times the cell size. Furthermore, VGG architecture family has been quite obsolete and . Thus, we updated the backbone with a variant of the state-of-the-art ConvNeXt [24]. ConvNeXt was recently introduced as a modernized version of ResNet [25] which is equipped with recent properties of the hierarchical vision transformer Swin [26]. Modifications were made on the macro scale, such as the change in the ratios between stages, the use of depthwise convolution in similar manner to ResNeXt [27], enlarging kernel size to 7, etc.; and on the micro scale, such as the switch from batch normalization [28] to layer normalization [29], from ReLU [30] to GELU [31], etc. These changes allow ConvNeXt to compete favorably with other architectures, including Swin, which is among the state-of-the-art for deep vision tasks.

We use an altered version of ConvNeXt-Tiny, which is most compact and efficient ConvNeXt type. ConvNeXt-Tiny starts with the stem block that reduce the input image to a feature map 1/4th the original size. The stem block is then followed by has four downsampling stages, each of which ends with a compression convolutional layer with reduction factor of 2. Thus we use a ConvNeXt-Tiny, which is truncated at before the compression layer of the 3rd stage, as the new backbone. The output of the truncated model reduces would have the size of 96x96x384. The modified ConvNeXt model is concatenated with two additional convolutional layers to reduce the feature maps to size 31x31x256. With this design, the receptive field is increased to 1240, which is almost the size of the input (1536).

We take a step further by adding three vision transformer encoder blocks [32], [33] to allow for global aggregation of information. Originally, the input of transformer encoder are flattened patches of images, which are added with positional embeddings to convey location information of each patch to the model. In LS-Netv2, the image patches are replaced with 1x1x256 feature segments from the 31x31x256 feature maps. This design is similar to the object detector DETR [34], where a CNN is used to produce a compact feature map first then a transformer is used in the later stage. The CNN provide inductive bias that allow the model to be trained

easier with smaller data. A pure transformer model would have more generalizability [?] but would rely on huge amount of data, which we do not have in our case, to achieve high performance. The benefit of this design will be shown using ablation study.

LS-Netv2 has the capability to detect multiple lines per cell by performing fixed number of inferences N, which is chosen to be larger than the maximum number of line segments exists in the cell. From our experiments, 10 is a safe choice. The ground truth  $y$  of each cell, which contains  $m$  actual line segments, is also perceived to contain N line segments  $y = \{y_i\}_{i=1}^N$ . However, the ground truth is now padded with  $N-m$  negative classification. Inspired by DETR [34], we line segment detection within each cell as a direct set prediction problem. In our loss computation step, we include an optimal bipartite matching between the N predictions an N ground truths.

Let  $\hat{y} = \{\hat{y}_i\}_{i=1}^N$  be the set of N predictions. The optimal bipartite matching between the N predictions an N ground truths, which is done using Hungarian algorithm, produces a permutation  $\hat{\sigma} \in \mathfrak{S}$  that satisfies:

$$\hat{\sigma} = \arg \min_{\sigma \in \mathfrak{S}} \sum_i^N \mathcal{L}_{match}(y_i, \hat{y}_{\sigma(i)}) \quad (1)$$

where  $\mathcal{L}_{match}(y_i, \hat{y}_{\sigma(i)})$  is the loss when pairing  $y_i$  and  $\hat{y}_{\sigma(i)}$ , which is reordered via permutation  $\sigma(i)$ . The loss considers both the classification loss and regression loss. Similarly to [34], this loss is defined as  $-\hat{p}_{\sigma(i)}(c_i) + \mathbb{1}_{\{c_i=\text{positive}\}} \mathcal{L}_{reg}(b_i, \hat{b}_{\sigma(i)})$ , where  $b_i$  and  $\hat{b}_{\sigma(i)}$  are line segment endpoint ground truth and permuted prediction at index i respectively.  $\mathcal{L}_{reg}(\cdot)$  is the regression loss and  $\hat{p}_{\sigma(i)}(c_i)$  is the predicted probability of class  $c_i$ , which is the classification ground truth at index i.

We propose a new version of LS-Net, called LS-Netv2. This version is updated with recent advancement in the deep learning space. Specifically, the feature extractor is altered with the state-of-the-art ConvNext model. Furthermore, inspired by DETR [34]. LS-Netv2 possess the ability to detect and locate multiple line-segments within each grid cell via multi-guess bipartite matching. Also, we propose the addition of a context module built from multi-head attentions layers, whose effectiveness in extracting global information has been proven. Finally, adjustments are made to the losses so that the training is more stable and requires less time to achieve plateau performance.

#### IV. EXPERIMENTS

The procedure and results of experiments are presented. Comparisons to state-of-the-art are also conducted.

##### A. Datasets

1) *Mendeley Powerline Dataset*: The Mendeley Powerline dataset comprises of 400 images of size 512x512, among which 200 contain powerlines while the remaining only contain background. This dataset possesses pixel level annotation. To adapt the annotation to LS-Netv2, pixel annotation of

individual powerlines are separated via connected component analysis. Then, a skeletonization algorithm, introduced at [35], is used to produce polylines tracing the powerlines. The polylines are simplified using RamerDouglasPeucker algorithm [36] and imposed on the 31x31 grid to produce the annotation for LS-Netv2.

*skeleton-tracing* [35], RamerDouglasPeucker algorithm [36]

2) *PLD-UAV*: PLD-UAV [37] contains two datasets of powerlines: power line dataset of urban scene (PLDU) and power line dataset of mountain scene (PLDM). In these datasets, the background are urban and mountain scene are quite cluttered and complex, however, the powerlines to be detected are still relatively observable. In this dataset, the boundary of powerlines are annotated at the pixel level. To adapt to LS-Netv2, we detected individual boundaries by dilating the pixel annotations and clustering the pixels via connected component analysis. From each cluster of pixels, a polygon is approximated and filled up with white pixels (positive labels). Then, skeletonization is used to produce the powerline tracing polylines, which can be imposed onto the 31x31 grid to produce the annotation for LS-Netv2. Manual inspection was done to make sure the procedure produce accurate labels. PLDU contains 453 training data points and 120 testing data points. PLDM contains 237 training data points and 50 testing data points.

3) *TTPLA*: TTPLA [38] is a newly introduced dataset. It consists of images taken from UAVs under varieties of conditions, such as scenes, angles, zooming levels and lighting conditions. Many instances in this dataset are imposed to problems like occlusion and blending, which make the detection more challenging. The annotation of TTPLA is at instance segmentation level via polygons precisely wrapping around power lines. TTPLA also possesses many data points that have powerlines being close to each other and powerline crossings, **which will prove the effectiveness of the new multi-line-segment capability of LS-Netv2**. To adapt it to LS-Netv2, similarly to the datasets above, we performed a image processing procedures which includes skeletonizing the blobs that fills the polygons annotations to generate polylines, which are simplified using RDP algorithim and then imposed on the 31x31 overlapping grid to find intersections, which are the labels for LS-Netv2. This dataset contains 1242 images, we uses 992 for training and 250 for testing. Examples are shown in Fig. ??.

4) *Proprietary dataset*: We also evaluated the method using a dataset of powerlines aggregated by EsmartSystems. This dataset is made from UAVs images taken by multiple clients of EsmartSystems, and thus, contains significant diversity in terms of scenes, angles, zooming levels, weather and lighting conditions. This dataset has polyline annotations tracing the trajectories of the powerlines. The polyline annotation can be imposed onto the 31x31 grid to produce the annotation for LS-Netv2. In this dataset, the lines annotated includes conductors, guy wires and overhead ground wires. Conductors are normally in parallel. However, the other two type of wires can have different directions and, hence, there are many visual crossings. Included in the dataset are also some instances of line segments being in close proximity. The dataset contains

2961 images for training and 468 images for testing. Examples are shown in Fig. ??.

## B. Implementation Details

### C. Evaluation Metrics

## V. CONCLUSION

The conclusion goes here.

## APPENDIX A PROOF OF THE FIRST ZONKLAR EQUATION

Appendix one text goes here.

## APPENDIX B

Appendix two text goes here.

## ACKNOWLEDGMENT

The authors would like to thank... [?]

## REFERENCES

- [1] R. Kasturi and O. I. Camps, "Wire detection algorithms for navigation," 2002.
- [2] C. Steger, "An unbiased detector of curvilinear structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 2, pp. 113–125, 1998.
- [3] G. Yan, C. Li, G. Zhou, W. Zhang, and X. Li, "Automatic extraction of power lines from aerial images," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 3, pp. 387–391, 2007.
- [4] Z. Li, Y. Liu, R. Walker, R. Hayward, and J. Zhang, "Towards automatic power line detection for a uav surveillance system using pulse coupled neural filter and an improved hough transform," *Machine Vision and Applications*, vol. 21, no. 5, pp. 677–686, Aug 2010. [Online]. Available: <https://doi.org/10.1007/s00138-009-0206-y>
- [5] J. Candamo, R. Kasturi, D. Goldgof, and S. Sarkar, "Detection of thin lines using low-quality video from low-altitude aircraft in urban settings," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 3, pp. 937–949, 2009.
- [6] I. Golightly and D. Jones, "Visual control of an unmanned aerial vehicle for power line inspection," in *ICAR '05. Proceedings., 12th International Conference on Advanced Robotics*, 2005., 2005, pp. 288–295.
- [7] Z. Li, Y. Liu, R. Hayward, J. Zhang, and J. Cai, "Knowledge-based power line detection for uav surveillance and inspection systems," in *2008 23rd International Conference Image and Vision Computing New Zealand*, 2008, pp. 1–6.
- [8] B. Alpatov, P. Babayan, and N. Shubin, "Power line detection using integrated vector radon transform," in *2016 5th Mediterranean Conference on Embedded Computing (MECO)*, 2016, pp. 162–164.
- [9] B. Song and X. Li, "Power line detection from optical images," *Neurocomputing*, vol. 129, pp. 350–361, 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231213009429>
- [10] C. Pan, X. Cao, and D. Wu, "Power line detection via background noise removal," in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2016, pp. 871–875.
- [11] J. Gubbi, A. Varghese, and P. Balamuralidhar, "A new deep learning architecture for detection of long linear infrastructure," in *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, 2017, pp. 207–210.
- [12] R. Grompone von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "Lsd: A fast line segment detector with a false detection control," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 4, pp. 722–732, 2010.
- [13] R. Madaan, D. Maturana, and S. Scherer, "Wire detection using synthetic data and dilated convolutional networks for unmanned aerial vehicles," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 3487–3494.
- [14] H. Zhang, W. Yang, H. Yu, H. Zhang, and G.-S. Xia, "Detecting power lines in uav images with convolutional features and structured constraints," *Remote Sensing*, vol. 11, no. 11, 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/11/1342>

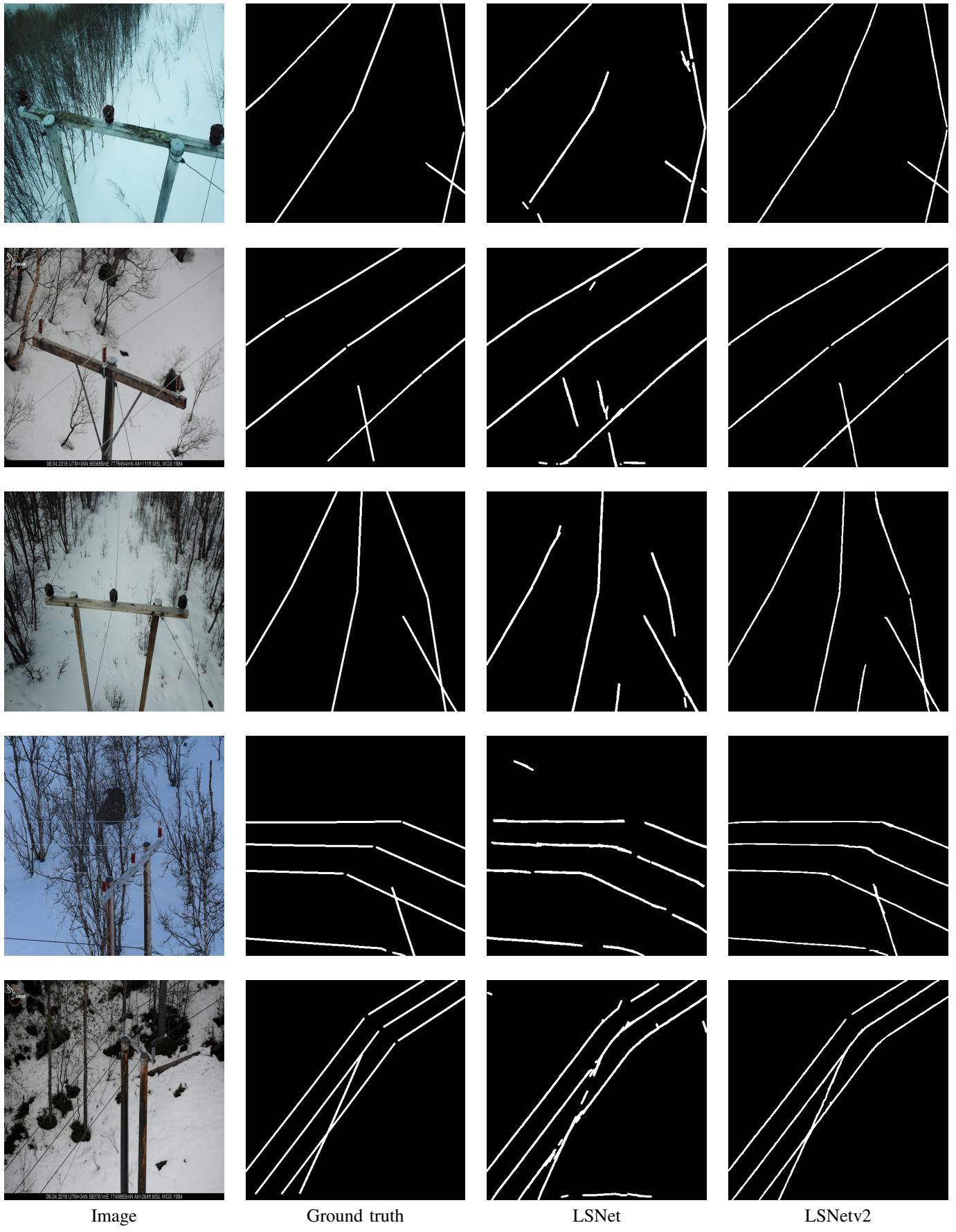


Fig. 1: 4 x 4



Fig. 2: 4 x 4



Fig. 3: 4 x 4

- [15] Y. Li, Z. Xiao, X. Zhen, and X. Cao, "Attentional information fusion networks for cross-scene power line detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 10, pp. 1635–1639, 2019.
- [16] E. Yetgin, B. Benligiray, and N. Gerek, "Power line recognition from aerial images with deep learning," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 5, pp. 2241–2252, 2019.
- [17] R. Abdelfattah, X. Wang, and S. Wang, "Plgan: Generative adversarial networks for power-line segmentation in aerial images," 2022. [Online]. Available: <https://arxiv.org/abs/2204.07243>
- [18] R. Jaffari, M. A. Hashmani, and C. C. Reyes-Aldasoro, "A novel focal phi loss for power line segmentation with auxiliary classifier u-net," *Sensors*, vol. 21, no. 8, 2021. [Online]. Available: <https://www.mdpi.com/1424-8220/21/8/2803>
- [19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [20] H. Choi, G. Koo, B. J. Kim, and S. W. Kim, "Weakly supervised power line detection algorithm using a recursive noisy label update with refined broken line segments," *Expert Systems with Applications*, vol. 165, p. 113895, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417420306953>
- [21] M. Bojarski, A. Choromanska, K. Choromanski, B. Firner, L. J. Ackel, U. Muller, P. Yeres, and K. Zieba, "Visualbackprop: Efficient visualization of cnns for autonomous driving," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 4701–4708.
- [22] S. J. Lee, J. P. Yun, H. Choi, W. Kwon, G. Koo, and S. W. Kim, "Weakly supervised learning with convolutional neural networks for power line localization," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017, pp. 1–8.
- [23] V. N. Nguyen, R. Jenssen, and D. Roverso, "Ls-net: fast single-shot line-segment detector," *Machine Vision and Applications*, vol. 32, no. 1, p. 12, Oct 2020. [Online]. Available: <https://doi.org/10.1007/s00138-020-01138-6>
- [24] Z. Liu, H. Mao, C. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11966–11976, 2022.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2015.
- [26] 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," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9992–10002, 2021.
- [27] S. Xie, R. B. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5987–5995, 2017.
- [28] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *ArXiv*, vol. abs/1502.03167, 2015.
- [29] J. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *ArXiv*, vol. abs/1607.06450, 2016.

- [30] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ser. ICML’10. Madison, WI, USA: Omnipress, 2010, p. 807814.
- [31] D. Hendrycks and K. Gimpel, “Bridging nonlinearities and stochastic regularizers with gaussian error linear units,” *ArXiv*, vol. abs/1606.08415, 2016.
- [32] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” *ArXiv*, vol. abs/2010.11929, 2020.
- [33] A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” *ArXiv*, vol. abs/1706.03762, 2017.
- [34] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, “End-to-end object detection with transformers,” *ArXiv*, vol. abs/2005.12872, 2020.
- [35] L. Huang, “Skeleton tracing,” <https://github.com/LingDong-/skeleton-tracing>, 2021.
- [36] D. H. Douglas and T. K. Peucker, “Algorithms for the reduction of the number of points required to represent a digitized line or its caricature,” *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 10, pp. 112–122, 1973.
- [37] H. S., “Pld-uav,” <https://github.com/SnorkerHeng/PLD-UAV>, 2019.
- [38] R. Abdelfattah, X. Wang, and S. Wang, “Ttpla: An aerial-image dataset for detection and segmentation of transmission towers and power lines,” in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, November 2020.



**Michael Shell** Biography text here.

**John Doe** Biography text here.

**Jane Doe** Biography text here.