

---

# Synthetic Imagery Aided Geographic Domain Adaptation for Rare Energy Infrastructure Detection in Remotely Sensed Imagery

---

**Wei Hu, Tyler Feldman, Eddy Lin, Jose Moscoso, Yanchen J. Ou,  
Natalie Tarn, Baoyan Ye, Wendy Zhang, Jordan M. Malof, Kyle Bradbury**

Duke University

{wei.hu, tyler.feldman, eddy.lin, jose.moscoso, yanchen.ou, natalie.tarn,  
baoyan.ye, wendy.zhang, jordan.malof, kyle.bradbury}@duke.edu

## Abstract

Object detection in remotely sensed data is frequently stymied by applications in geographies that are different from that of the training data. When objects are rare, the problem is exacerbated further. This is true of assessments of energy infrastructure such as generation, transmission, and end-use consumption; key to electrification planning as well as for effective assessment of natural disaster impacts which are varying in frequency and intensity due to climate change. We propose an approach to domain adaptation that requires only unlabeled samples from the target domain and generates synthetic data to augment training data for targeted domain adaptation. This approach is shown to work consistently across four geographically diverse domains, improving object detection average precision by 15.7% on average for small sample sizes.

## 1 Introduction

Developing effective climate change mitigation and adaptation strategies requires an accurate and up-to-date understanding of climate change drivers, especially anthropogenic greenhouse gas emissions (GHG) from the energy sector [1]. Energy systems, and the electricity sector in particular, are changing rapidly: increasing renewable energy globally [2], in particular, the rise of many forms of distributed energy, and grid expansion in Sub-Saharan Africa and parts of southeast Asia [3]. Understanding these developments and their GHG implications in near real-time is important for establishing realistic strategies and policies for addressing climate change impacts.

However, these important energy systems data are often challenging to obtain and keep up-to-date, and vary in availability from region-to-region. One potential solution to this limitation is to use remote sensing data, and numerous studies have demonstrated that computer vision techniques, such as object detection and image segmentation are capable of accurately detecting energy infrastructure in remotely sensed data [4, 5, 6]. These techniques have been shown to work well in situations when the training data are very representative of the validation data. In real applications of these techniques, we often do not have labeled training data for the region to which we wish to apply our models: in such cases, we need to adapt our approach to the new domain (i.e. domain adaptation or shift).

Recent evidence on domain adaptation [7, 8, 9] suggests that deep learning models do not generalize well to imagery from new geographic locations. This problem arises due to the visual variability of overhead imagery due to atmospheric conditions, sensor characteristics, regional flora, and differences in the built environment. Some of these differences may cause the model to generalize unpredictably, often very poorly [9].

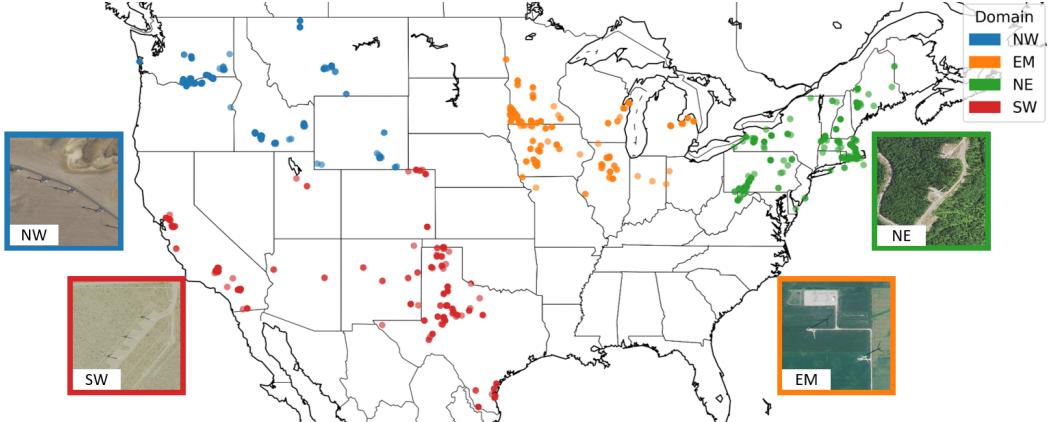


Figure 1: Real imagery source locations, geographic domains, and examples from each geographic domain.

While there has been some progress on domain adaptation for deep learning [10, 11], this remains challenging especially for large train/test data domain gaps [12] and we remain far from being able to train a model once with limited training data and apply it globally, a major hindrance for autonomous mapping. Recent work on synthetic imagery generation for training augmentation for rare objects offers a potential solution [9, 13, 14, 15]. These works showed that synthetic training imagery could supplement training sets, resulting in performance improvements. In this work we build upon the recent SIMPL method [14] because it is replicable, easily adapted to new problems, and designed for object detection problems. In this work, we adapt their approach for geographic domain adaptation, and demonstrate its effectiveness for mapping energy infrastructure in satellite imagery across substantially larger, and more diverse, geographies.

The contributions of this paper are three-fold: (1) we created an overhead imagery dataset with labeled wind turbines from four carefully selected geographies for evaluating domain adaptation effectiveness; (2) the dataset includes training augmentation imagery with *synthetic* wind turbine objects using *unlabeled* data from target domains; (3) we evaluated performance on held-out data for all combinations of domains and demonstrate efficacy of the technique.

## 2 Methodology

This work focuses on object detection of wind turbines. While we could demonstrate this process for any type of energy infrastructure, we use wind turbines because, aside from size differences, they are relatively homogeneous in appearance, to reduce the number of variables for which we need to control. To investigate the use of synthetically generated imagery for domain adaptation for wind turbines, we will first create a baseline dataset of real wind turbines. Then, to create the synthetic imagery, we use imagery from the same dataset *without* wind turbines and augment it with *synthetic* wind turbine objects. Lastly, we will train object detection models to evaluate their performance across domains.

### 2.1 Dataset creation

We first created the real imagery dataset by sampling from the National Agricultural Imagery Program (NAIP[16]) dataset, which is 0.6 m/pixel resolution across the U.S. We samples from 4 geographic regions (i.e. domains) that we called Northwest (NW), Northeast (NE), Eastern Midwest (EM), and Southwest (SW), chosen to vary in the visual appearance (Figure 1). Using the U.S. Wind Turbine Database (USWTDB[17]) to identify locations of wind turbines locations and manually annotated those turbines within each each image. In total, 1,100 real images were collected in total and we split 400 into the training set, 400 into the validation set, and 300 without wind turbines for synthetic imagery creation, balanced equally across the four geographic domains.

Within each of the 4 domains (NW, NE, EM, and SW) there may be some internal variation, so to increase the homogeneity within each domain, we spatially clustered the data within each domain

and performed stratified sampling from those clusters each to ensure equal spatial representation. Using this approach, we sampled data for the baseline dataset and include 100 images in the training and 100 images in the validation set from each of the four domains. A small number of images were chosen to use in this study to simulate having few training examples, as may often be the case for rare objects, such as energy infrastructure.

Next, we generated the synthetic data to supplement the real training data. We do this by taking a real image without a wind turbine present and superimposed a 3D model of a wind turbine on top of that image using the software CityEngine as described in [14] and as demonstrated in Figure 2 (creating corresponding ground truth as well). Selecting the ratio of synthetic-to-real data to include for training was another consideration requiring evaluation. Experimentally, we swept over prospective values of the ratio and determined that 0.75:1 was optimal in this setting, generating 75 synthetic images for each of the geographic domains. For more information on dataset construction, see Appendix A. With real and synthetic imagery created, we test wind turbine detection performance with and without the addition of synthetic imagery across our different domains.

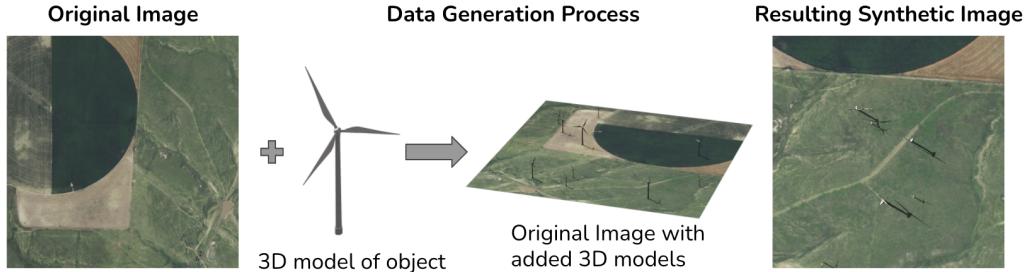


Figure 2: Synthetic imagery generation process.

The two final datasets for each domain (show in Figure 2) are:

- Baseline dataset: 100 real source domain training images, 100 target domain real testing images.
- Synthetic-augmented dataset: 100 real source domain training images *and 75 target domain synthetic training images*, target domain 100 real testing images.

## 2.2 Geographic domain adaptation experimental design

To test the efficacy of synthetic imagery for domain adaptation on our wind turbine dataset, we setup a series of experiments divided into two groups: within-domain and cross-domain experiments. For each experiment we define a source domain as the geographic region the training images are sourced from. This could be any of the four domains shown in Figure 2. The target domain is the region on which we validate our trained object detection model’s performance. For within-domain experiments, the source domain and the target domain are the same. Cross-domain experiments have different source and target domains. In all experiments, the validation data are distinct from the training data, regardless of source or target domain.

Our baseline experiments are comprised of training a YOLOv3 [18] object model on each of the pairs of the four domains in our dataset, resulting in 16 experiments, varying source and target domains. We then repeat these 16 experiments, except we augment the real training data from the source domain with synthetic samples from the target domain. We repeat every experiment four times to account for model variance. The hypothesis we are testing is that synthetic imagery will help our model adapt to unseen domains, yielding performance gains.

## 3 Results and discussion

Performance of models trained with and without added synthetic imagery show are shown in Table 1. Average precision (AP) is reported for all experiments and two summary metrics are also reported: the percentage improvement in AP and the percentage closure of domain gap (CDG). CDG considers

Table 1: Average Precision of all pair-wise domain experiments. The 95% confidence intervals shown are based on four runs of each experimental condition. In each experiment (one per row), the highest average precision is noted in **bold**

Source domain	Target domain	Baseline $\pm 2\sigma$	Adding synthetic $\pm 2\sigma$	Average improvement%	Average closure% of domain gap
EM	EM	0.822 $\pm$ 0.067	<b>0.919</b> $\pm$ 0.016	11.8%	-
NE		0.567 $\pm$ 0.019	<b>0.698</b> $\pm$ 0.038	23.1%	51.4%
NW		0.358 $\pm$ 0.061	<b>0.424</b> $\pm$ 0.114	18.4%	14.2%
SW		0.449 $\pm$ 0.160	<b>0.626</b> $\pm$ 0.180	39.4%	47.5%
EM	NE	0.387 $\pm$ 0.031	<b>0.487</b> $\pm$ 0.114	25.8%	23.5%
NE		0.812 $\pm$ 0.028	<b>0.842</b> $\pm$ 0.013	3.7%	-
NW		0.666 $\pm$ 0.061	<b>0.709</b> $\pm$ 0.049	6.5%	29.5%
SW		0.412 $\pm$ 0.045	<b>0.521</b> $\pm$ 0.089	26.5%	27.3%
EM	NW	0.485 $\pm$ 0.064	<b>0.521</b> $\pm$ 0.054	7.4%	8.8%
NE		0.746 $\pm$ 0.018	<b>0.770</b> $\pm$ 0.032	3.2%	16.1%
NW		0.895 $\pm$ 0.071	<b>0.915</b> $\pm$ 0.023	2.2%	-
SW		0.659 $\pm$ 0.111	<b>0.693</b> $\pm$ 0.066	5.2%	14.4%
EM	SW	0.093 $\pm$ 0.016	<b>0.113</b> $\pm$ 0.008	20.9%	4.1%
NE		0.121 $\pm$ 0.029	<b>0.134</b> $\pm$ 0.030	10.7%	2.9%
NW		0.149 $\pm$ 0.029	<b>0.197</b> $\pm$ 0.024	32.2%	11.5%
SW		0.566 $\pm$ 0.035	<b>0.568</b> $\pm$ 0.104	0.4%	-
Within-domain average		0.774 $\pm$ 0.050	<b>0.811</b> $\pm$ 0.039	4.8%	-
Cross-domain average		0.425 $\pm$ 0.054	<b>0.491</b> $\pm$ 0.067	15.7%	20.9%

the best case scenario to be the within-class setting, and the worst case to be the cross domain setting without synthetic data, this difference defines a "gap", and the amount of that gap which the addition of the synthetic data recovers is the percentage CDG (see Appendix D). Across all experimental settings, *all within-domain and cross-domain experiments*, adding synthetic imagery improved average precision over the baseline. Adding synthetic imagery yielded far greater performance gains in the cross-domain setting, when the target domain differed from the source domain. The average across domains is summarized at the bottom of Table 1 and shows while for within class experiments, adding synthetic data offered a nontrivial 4.8% improvement in average precision, adding targeted synthetic imagery in the cross-domain setting yielded a 15.7% improvement in average precision compared to the baseline and closed the domain gap by 20.9%.

These improvements in domain adaptation through synthetic data augmentation are based only on unlabeled training data from the target domain, which is often available. This approach is promising for overcoming domain gaps in energy infrastructure assessment and for other types of infrastructure. Further work is needed to explore other types of infrastructure to test the generalizability of this approach across object classes. These results are encouraging for the long-term goal of being able to apply energy infrastructure object detection models to a variety of unseen domains around the world wherever energy systems are changing.

## Acknowledgments and Disclosure of Funding

This work was supported by the Duke University Bass Connections program, the Duke University Data+ program, and the Duke University Energy Initiative. We thank Matt Robbins and Kate Chen for their contributions to this project.

## References

- [1] Intergovernmental Panel on Climate Change (IPCC). Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change. *Intergovernmental Panel on Climate Change (IPCC), Cambridge University Press, Cambridge*, 2021. In Press.
- [2] Linda Capuano. International energy outlook 2020, 2020.
- [3] International Energy Agency (IEA), International Renewable Energy Agency (IRENA) and United Nations Statistics Division (UNSD) and World Bank, and World Health Organization (WHO). World bank global electrification database: Access to electricity (% of population, 2020.
- [4] Jordan M Malof, Kyle Bradbury, Leslie M Collins, and Richard G Newell. Automatic detection of solar photovoltaic arrays in high resolution aerial imagery. *Applied energy*, 183:229–240, 2016.
- [5] Xinyu Liu, Hao Jiang, Jing Chen, Junjie Chen, Shengbin Zhuang, and Xiren Miao. Insulator detection in aerial images based on faster regions with convolutional neural network. In *2018 IEEE 14th International Conference on Control and Automation (ICCA)*, pages 1082–1086. IEEE, 2018.
- [6] Ashley Varghese, Jayavardhana Gubbi, Hrishikesh Sharma, and P Balamuralidhar. Power infrastructure monitoring and damage detection using drone captured images. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 1681–1687. IEEE, 2017.
- [7] Onur Tasar, Alain Giros, Yuliya Tarabalka, Pierre Alliez, and Sébastien Clerc. Daugnet: Unsupervised, multisource, multitarget, and life-long domain adaptation for semantic segmentation of satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(2):1067–1081, 2020.
- [8] Onur Tasar, Yuliya Tarabalka, Alain Giros, Pierre Alliez, and Sébastien Clerc. Standardgan: Multi-source domain adaptation for semantic segmentation of very high resolution satellite images by data standardization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 192–193, 2020.
- [9] Fanjie Kong, Bohao Huang, Kyle Bradbury, and Jordan Malof. The synthinel-1 dataset: a collection of high resolution synthetic overhead imagery for building segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1814–1823, 2020.
- [10] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153, 2018.
- [11] Vishal M Patel, Raghuraman Gopalan, Ruonan Li, and Rama Chellappa. Visual domain adaptation: A survey of recent advances. *IEEE signal processing magazine*, 32(3):53–69, 2015.
- [12] Yang Zou, Zhiding Yu, BVK Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European conference on computer vision (ECCV)*, pages 289–305, 2018.
- [13] Jacob Shermeyer, Thomas Hossler, Adam Van Etten, Daniel Hogan, Ryan Lewis, and Daeil Kim. Rareplanes: Synthetic data takes flight. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 207–217, 2021.
- [14] Yang Xu, Bohao Huang, Xiong Luo, Kyle Bradbury, and Jordan M Malof. Simpl: Generating synthetic overhead imagery to address zero-shot and few-shot detection problems. *arXiv preprint arXiv:2106.15681*, 2021.
- [15] Sanghui Han, Alex Fafard, John Kerekes, Michael Gartley, Emmett Ientilucci, Andreas Savakis, Charles Law, Jason Parhan, Matt Turek, Keith Fieldhouse, et al. Efficient generation of image chips for training deep learning algorithms. In *Automatic Target Recognition XXVII*, volume 10202, page 1020203. International Society for Optics and Photonics, 2017.

- [16] U.S. Department of Agriculture. National agricultural imagery program, 2003-present.
- [17] United States Geological Survey (USGS), Lawrence Berkeley National Laboratory (LBNL), and American Clean Power Association (ACP). The u.s. wind turbine database, 2016-present.
- [18] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

## Appendix A Synthetic data generation

Several considerations were taken in the design of the synthetic imagery creation to make synthetic imagery as similar to real imagery as possible. Firstly, we modeled the size of synthetic turbines based on the distribution of sizes of the real turbines. Second, since the camera angle is not always perpendicular to the ground in overhead imagery, when taking snapshots of synthetic 3D scenes we distributed the camera angle from 60 to 90 degrees, with 50% of the images being taken at 90 degrees.

When selecting the background images for the synthetic data generation, we chose images at random, but geographically close to the real images in the validation dataset so that the synthetic imagery would look similar to the target domain.

Another consideration was the relative quantity of real and synthetic training data. To figure out what amount of synthetic imagery would yield the largest gain in performance, we ran experiments to determine the optimal ratio. If we add too much synthetic data, we run the risk of overfitting to synthetic data; if we add too little, then it will have little impact on performance. To find the optimal real:synthetic ratio, we designed an experiment where we train YOLOv3 object detection models using 1:0, 1:0.5, 1:0.75, 1:1, 1:2 real to synthetic ratios. We then evaluated these models on the same testing set. 1:0.75 yielded the best performance of the options investigated, so we designed our experiments using the 1:0.75 real-to-synthetic ratio.

## Appendix B Within-domain and cross-domain experimental designs

To further clarify our within-domain and cross-domain experimental designs, we summarize those in Figure 3. In the within-domain setting, the source and target domains are the same, while in the cross-domain setting, the source and target domains are different. In all cases, separate data are used for validation that are distinct from the training data.

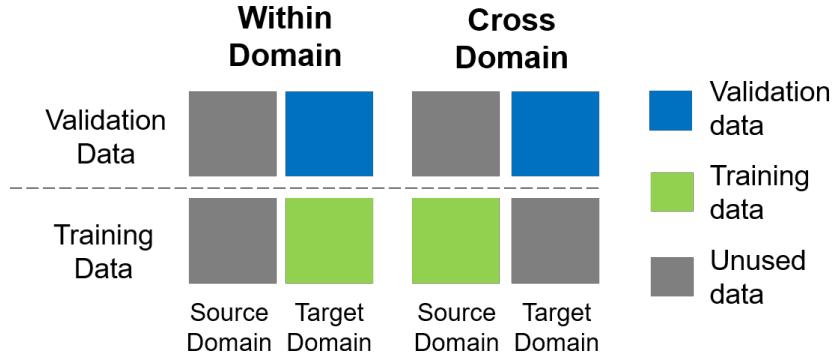


Figure 3: Within-domain and cross-domain experimental design.

## Appendix C Experimental setup

We trained object detection models using the YOLOv3 framework implemented by Ultralytics<sup>1</sup>. Each model was trained for 300 epochs using a batch size of 10. The input image pixel size is 608\*608. We evaluate model performance using average precision (AP). Each experiment was repeated 4 times to reduce variance and the average AP values were calculated at Intersection over Union (IoU) threshold of 0.5. All model training and testing were conducted using an NVIDIA GeForce RTX 2080 Ti GPU. Code base and instructions for reproduction of this paper are found at [https://figshare.com/articles/dataset/Synthetic\\_Wind\\_Turbine\\_Dataset/16639546](https://figshare.com/articles/dataset/Synthetic_Wind_Turbine_Dataset/16639546).

<sup>1</sup><https://github.com/ultralytics/yolov3>

## Appendix D Percentage closure of domain gap metric

The percentage closure of domain gap (CDG) considers the best case scenario to be the within-class setting, and the worst case to be the cross domain setting without synthetic data, this difference defines a "gap", and the amount of that gap which the addition of the synthetic data recovers is the percentage CDG. If average precision of within-class setting is  $AP_w$ , the average precision of cross-domain experiment without synthetic data is  $AP_c$ , and the average precision of the cross domain experiment with synthetic data from the target domain added is  $AP_{c+s}$ , then the percentage closure of the domain gap is  $CDG$ , given as follows:

$$CDG = \frac{AP_{c+s} - AP_c}{AP_w - AP_c} \times 100\%$$

## Appendix E Results for all within-domain and cross-domain experiments

The data from Table 1 was based on four repeated runs of each experimental pair for both the case of the baseline and experimental (adding synthetic data) conditions. The full results of each experiment for the baseline and experimental conditions are shown for the eastern midwest (EM) domain in Table 2, for the northeast (NE) domain in Table 3, for the northwest (NW) domain in Table 4, and for the southwest (SW) domain in Table 5.

Table 2: Average Precision of all experiments where source domain is EM.

Source domain	Target domain	Baseline AP	Adding Synthetic AP
EM	EM	0.804	0.929
EM	EM	0.862	0.912
EM	EM	0.786	0.913
EM	EM	0.834	0.923
EM	NE	0.373	0.552
EM	NE	0.375	0.462
EM	NE	0.399	0.422
EM	NE	0.402	0.513
EM	NW	0.527	0.533
EM	NW	0.486	0.542
EM	NW	0.479	0.519
EM	NW	0.449	0.481
EM	SW	0.104	0.132
EM	SW	0.090	0.104
EM	SW	0.085	0.098
EM	SW	0.095	0.119

Table 3: Average Precision of all experiments where source domain is NE.

Source domain	Target domain	Baseline AP	Adding Synthetic AP
NE	EM	0.553	0.680
NE	EM	0.573	0.701
NE	EM	0.570	0.687
NE	EM	0.573	0.723
NE	NE	0.796	0.839
NE	NE	0.808	0.835
NE	NE	0.817	0.847
NE	NE	0.829	0.848
NE	NW	0.752	0.786
NE	NW	0.764	0.781
NE	NW	0.746	0.758
NE	NW	0.744	0.754
NE	SW	0.129	0.135
NE	SW	0.100	0.139
NE	SW	0.122	0.130
NE	SW	0.132	0.132

Table 4: Average Precision of all experiments where source domain is NW.

Source domain	Target domain	Baseline AP	Adding Synthetic AP
NW	EM	0.370	0.447
NW	EM	0.383	0.339
NW	EM	0.366	0.453
NW	EM	0.314	0.457
NW	NE	0.623	0.724
NW	NE	0.670	0.734
NW	NE	0.693	0.679
NW	NE	0.679	0.700
NW	NW	0.911	0.905
NW	NW	0.842	0.905
NW	NW	0.911	0.926
NW	NW	0.916	0.924
NW	SW	0.139	0.198
NW	SW	0.169	0.187
NW	SW	0.138	0.214
NW	SW	0.151	0.190

Table 5: Average Precision of all experiments where source domain is SW.

Source domain	Target domain	Baseline AP	Adding Synthetic AP
SW	EM	0.359	0.672
SW	EM	0.413	0.591
SW	EM	0.483	0.518
SW	EM	0.542	0.722
SW	NE	0.402	0.531
SW	NE	0.445	0.456
SW	NE	0.403	0.546
SW	NE	0.396	0.552
SW	NW	0.668	0.648
SW	NW	0.695	0.686
SW	NW	0.578	0.722
SW	NW	0.695	0.713
SW	SW	0.581	0.491
SW	SW	0.542	0.586
SW	SW	0.563	0.605
SW	SW	0.576	0.590