Study of Zero-shot Learning for Visual Search on Satellite and Aerial Images

A. Chuong Dang, Ion-George Todoran, Srushti Rashmi Shirish

June 26th – 30th , 2022

Presenter: A. Chuong Dang [DC]

Title: Machine Learning Engineer

Email: dang.anh.chuong@woven-planet.global

A. Chuong Dang [a.k.a: DC]

Email: dan.anh.chuong@woven-planet.global

Education:

- Bachelor in Mechanical Engineering, Tohoku University, JAPAN.
- Master Degree in Information Science, Tohoku University, JAPAN.

Profession:

- Machine Learning Engineer at Automated Mapping Platform, Woven Alpha Inc., Woven Planet Holdings Inc. Member of Machine Learning Platform team, Unified Pipeline.

Interest:

- Development/Applications of Machine Learning, Deep Learning algorithms.

amp

Motivation

- Top-down images are semantically complex.
- Satellite image source is abundant, but low utilization.
- Difficulty in efficient data managing.

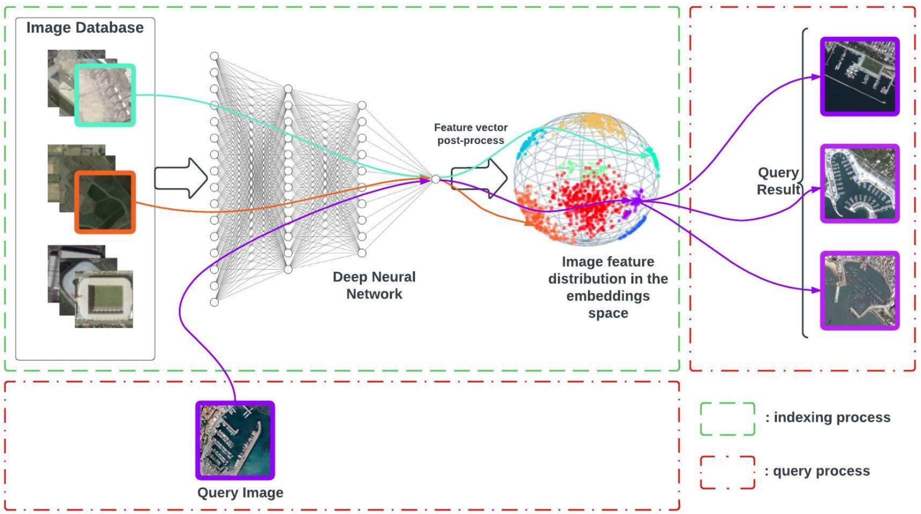
woven planet

Confidential

Research Summary

In this study, we:

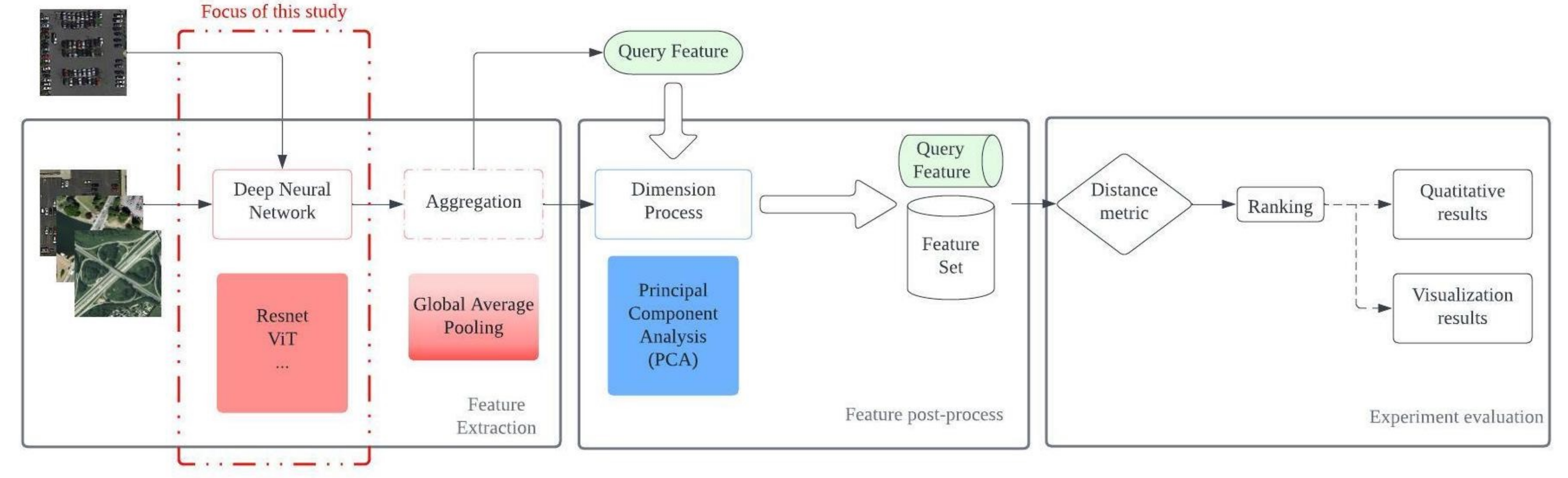
- Present a Visual Search system based on latest Deep Learning techniques.
- Propose to mitigate diverse categories issue by *"zero-shot learning"* method.
- Introduce to improve system performance by pre-training feature embedding model using top-down images.
- Study the possibility of applying unsupervised method to alleviate the problem of lacking labeled data.



woven planet

Confidential

Approach



Research questions:

- How the data effects system performance?
- How important is the role of feature extractor?
- How to utilize data source better?

Experimental Settings:

- Datasets: UC Merged Land Use, AID, RESISC45.
- DNNs: ResNet, Vision Transformers (ViT).
- Training methods: supervised and unsupervised.

woven planet

Results I

Confidential

Supervised pre-trained using photographic versus aerial imagery datasets

Test dataset	Pre-trained dataset	mAP	R@1	Pre-trained dataset	mAP	R@1
UC Merged Land Use	ImageNet1k	***58.9***	***92.9***	AID (x224)	60.0	91.0
	Places365	54.3	90.2	AID (x320)	62.7	90.9
	ImageNet1k & Place365	57.5	92.6	RESISC45	***78.6***	***95.9***
AID	ImageNet1k	***44.6***	***85.4***			
	Places365	42.3	83.3	RESISC45	***69.3***	***89.2***
	ImageNet1k & Place365	44.4	84.0			
RESISC45	ImageNet1k	34.0	78.7	AID (x224)	***44.0***	***80.9***
	Places365	33.2	77.9			
	ImageNet1k & Place365	***35.0***	***80.3***	AID (x320)	43.4	79.6

- Pre-trained on aerial imagery datasets have a positive effect on system performance.
- Pre-trained using unsupervised method?

Results II

Unsupervised pre-trained using aerial imagery datasets

Confidential

- Improved/comparative system performance yet may not surpass supervised methods.
- Help to utilizing large amount of unlabeled data.
- Only helpful when having access to a decent amount of data.

Test dataset	Pre-trained dataset	mAP	R@1
UC Merged Land Use	ImageNet1k	58.9	*94.7*
	AID	55.0	93.1
	RESISC45	*63.0*	93.8
AID	ImageNet1k	46.7	88.6
	RESISC45	*52.1*	*90.1*
RESISC45	ImageNet1k	*36.6*	*84.6*
	AID	36.1	84.0

woven planet

Results III

Vision Transformer (ViT) as feature extractor

Test dataset	Backbone architecture	Pre-trained dataset	Pre-trained method	mAP	R@1
UC Merged Land Use	ResNet50	ImageNet1k	Unsuppervised	58.9	94.7
	ViT-S/16			63.3	***95.7***
	ViT-S/8			***67.0***	95.4
AID	ResNet50			46.7	88.6
	ViT-S/16			49.8	90.2
	ViT-S/8			***53.7***	***91.7***
RESISC45	ResNet50			36.6	84.6
	ViT-S/16			39.7	86.9
	ViT-S/8			***43.0***	***88.8***

- Utilizing latest DNNs architecture improved performance of the system by a good margin.
- Still existing challenges and drawbacks.

Results IV

Confidential

Ablation Study: Impact of removing feature dimensionality reduction

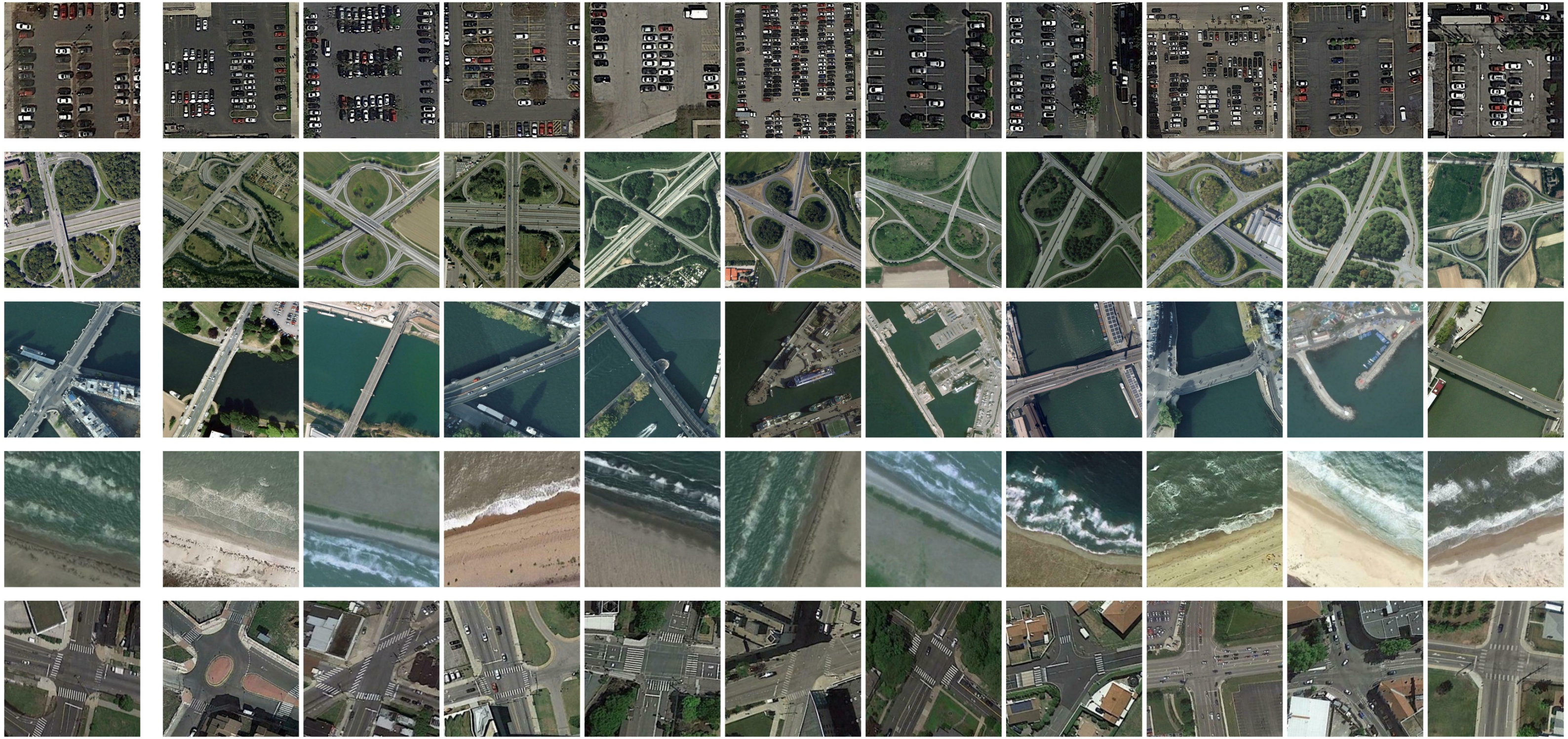
Test dataset	Dimension Reduction Method	Test dataset	Pre-trained dataset	Pre-trained method	mAP	mAP drop ↓	R@1	R@1 drop ↓
ResNet50	PCA	AID	ImageNet1k	Unsuppervised	46.7	5.1 ↓	88.6	0.6 ↓
	None				41.6		88.0	
ViT-S/16	PCA				49.8	4.5 ↓	90.2	0.3 ↓
	None				45.3		88.9	
ViT-S/8	PCA				53.7	4.8 ↓	91.7	1.1 ↓
	None				48.9		90.6	

- Removing dimension reduction method yields negative impact on system’s performance.
- Necessity of using dimension reduction method in case which requires high accuracy.
- Yet, dimension reduction is not scalable → further research!

woven planet

Results V

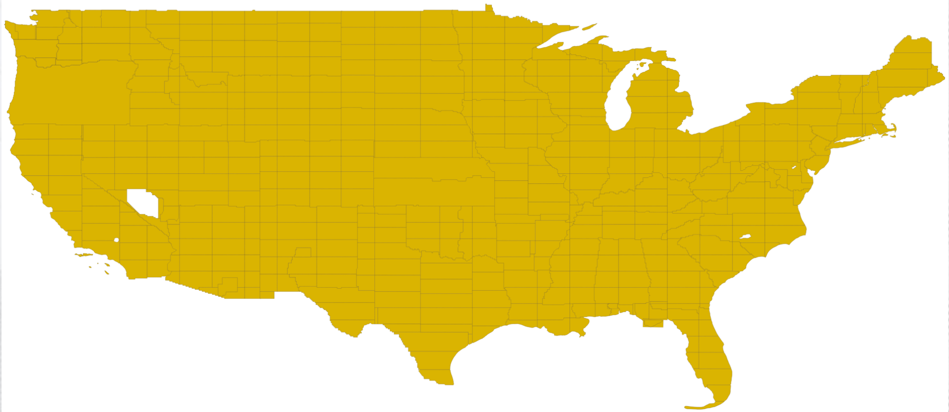
Visualization results



Challenges and next steps

Challenges (further study):
- Further experiment with ViTs and unsupervised training.
- Research towards scalable dimension reduction method.
- Improve system performance evaluation process.

Next steps (application investigation):
- Define tiling strategy.
- Indexing at large scale.
- API for query images.



woven planet

Thank you!

References

1. M. Tarasiou and S. Zafeiriou, “DeepSatData: Building large scale datasets of satellite images for training machine learning models,” CoRR, abs/2104.13824, 2021.
2. S.Bell and K. Bala, “Learning visual similarity for product design with convolutional neural networks,” ACM Trans. Graph., Vol. 34, No. 4, pp 1–10, 2015.
3. K. Sohn, “Improved Deep Metric Learning with Multi-class N-pair Loss Objective,” NIPS, 2016.
4. R. Keisler, S. Skillman, S. Gonnabathula, J. Poehnelt, X. Rudelis, and M. Warren, “Visual search over billions of aerial and satellite images,” Comput. Vis. Image Underst. vol. 187, Issue C, pp 1-6, 2019.
5. The SpaceNet Partners, “SpaceNet5: Automated Road Network Extraction and Route Travel Time Estimation from Satellite Imagery,” https://spacenet.ai/sn5-challenge/, Accessed May 5th 2022.
6. J. Johnson, M. Douze, and H. JÅLegou, “Billion-scale similarity search with GPUs,” IEEE Transactions on Big Data, vol. 7, No. 3, pp. 535-547, 2019.
7. C. Wengert, M. Douze, and H. JÅLegou, “Bag-of-colors for improved image search,” In ACM Multimedia, pp.1437–1440, 2011.
8. M. Park, J. Jin, and L. Wilson, “Fast content-based image retrieval using quasi-gabor filter and reduction of image feature dimension,“ Fifth IEEE Southwest Symposium on Image Analysis and Interpretation, 2002.
9. R. Arandjelovic and A. Zisserman, “Three things everyone should know to improve object retrieval,“ CVPR, 2012.
10. A. Krizhevsky and G. Hinton, “Using very deep autoencoders for content-based image retrieval,” in Proceedings of the European Symposium of Artifical Neural Networks (ESANN), 2011.
11. J. Ng, F. Yang, and L. Davis, “Exploiting local features from deep networks for image retrieval,” CVPR Workshops, 2015.
12. G. Tolias, T. Jenicek, and O. Chum, “Learning and aggregating deep local descriptors for instance-level recognition,” ECCV, 2020.
13. D. Yi, Z. Lei, S. Liao, and S. Z. Li, ”Deep metric learning for person re-identification,” 22nd International Conference on Pattern Recognition, 2014.
14. W. Ge, W. Huang, D. Dong, and M.R. Scott, “Deep metric learning with hierarchical triplet loss,” ECCV, 2018.
15. M. Everingham, S.M.A. Eslami, L. Van Gool, C.K.I. Williams, J. Winn, and A. Zisserman, “The Pascal visual object classes challenge: A retrospective,” International Journal of Computer Vision, vol. 111, no. 1, pp. 98-136, 2015.
16. H. JÅLegou, M. Douze, and C. Schmid, “Product quantization for nearest neighbor search,” In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 1, pp. 117-128, 2011.
17. Y.Yang and S. Newsam, “Bag-Of-Visual-Words and Spatial Extensions for Land-Use Classification,” ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS), 2010.

References

18. G.-S. Xia, J.Hu, F. Hu, and B. Shi, “AID: A Benchmark Dataset for Performance Evaluation of Aerial Scene Classification,” In IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 7, pp. 3965-3981, 2017.
19. G. Cheng, J. Han, and X. Lu, “Remote Sensing Image Scene Classification: Benchmark and State of the Art,” In Proceedings of the IEEE, vol. 105, no. 10, pp. 1865-1883, 2017.
20. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. FeiFei, “Imagenet: A large-scale hierarchical image database,” CVPR, 2009.
21. B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, ”Places: A 10 million image database for scene recognition,” In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 6, pp. 1452-1464, 2018.
22. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” CVPR, 2016.
23. M. Caron, H. Touvron, I. Misra, H. JÅLegou, J. Mairal, P. Bojanowski, and A. Joulin, ”Emerging Properties in Self-Supervised Vision Transformers,” ICCV, 2021.
24. G. H. Golub and C. Reinsch, ”Singular Value Decomposition and Least Squares Solutions,” In: Bauer, F.L. (eds) Linear Algebra. Handbook for Automatic Computation, vol 2. Springer, 1971, pp.134-151.
25. Z. Zhong, L. Zheng, D. Cao, and S. Li, ”Re-ranking Person Reidentification with k-reciprocal Encoding,” CVPR, 2017.
26. K. Pearson, ”LIII. On lines and planes of closest fit to systems of points in space,” The London, Edinburgh, and Dublin philosophical magazine and journal of science, Series 6, vol. 2, Issue 11, pp.559-572, 1901.
27. A. Dosovitskiy, et al., ”An image is worth 16x16 words: Transformers for image recognition at scale,” ICLR, 2021.
28. K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick. ”Momentum Contrast for Unsupervised Visual Representation Learning,” CVPR, 2020.
29. X. Chen, H. Fan, R. Girshick, and K. He. ”Improved Baselines with Momentum Contrastive Learning,” CoRR abs/2003.04297, 2020.
30. X. Chen, S. Xie, and K. He. ”An Empirical Study of Training Self-Supervised Vision Transformers,” ICCV, 2021.
31. T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. ”A Simple Framework for Contrastive Learning of Visual Representations,” ICML, 2020.
32. J.B. Grill, et al., ”Bootstrap your own latent: A new approach to selfsupervised Learning,” NeurIPS, 2020.
33. X. Chen and K. He. ”Exploring Simple Siamese Representation Learning,” CVPR, 2021.