Similarity Metrics for Cartographic Sentinel-2 Multi-spectral Imagery Comparison

Algirdas Benetis, Vytautas Valaitis

Institute of Computer Science, Vilnius University

Vilnius University

Introduction

Drones localize through GPS signal transmission, but sometimes this is not possible due to interference or noise, and sensors alone are not enough for accurate positioning in the long run. In the era of digitization, many fields, including agriculture or the military industry, use drones for various purposes. Using an orthographic image similarity metric based on triplet neural networks is one way to determine the drone's location. The topic of this study is the calculation and comparison of similarity metrics based on the developed EfficientNet [10], EfficientNetV2 [11], MobileNet [7], ResNet [5] and VGG neural network architectures by using different band composites of Sentinel-2. The trained base layers of these networks are used in triplet neural networks. During the experiments of image processing time, distances from the anchor picture and precision metrics, which would help to determine more acceptable architectural configurations and band composites for comparing orthographic images and finding the location of drones, the results of the similarity metrics between band composites were compared with each other. By using combinations of bands, we can extract specific information from an image. E.g., there are combinations of bands that highlight geological, agricultural or vegetation features in an image. The final choice of the triplet neural network model and band combination may depend on various factors and, it is important to emphasize that it is worth considering all the results of the obtained metrics before applying the respective architectures to single cases, to evaluate the importance of each metric and composite in personalized application situations.

Experiments

An example of a photo triplet made up of orthographic photos. Anchor photo - left (2023), positive photo - middle (2019), negative photo - right (2019)

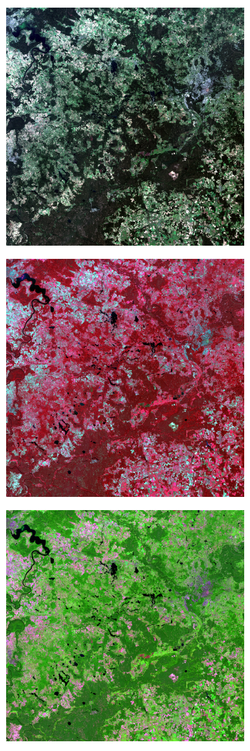
The differences of average distances between different architectures of triplet neural networks and their configurations in photo processing with True Color, False Color and Shor-wawe Infrared (SWIR) Sentinel-2 composites

True Color		
Network	Layer	Avg. d. diff.
EfficientNet-B2	NN(59) TR(0)	0,19
	NN(59) TR(25)	0,23
	NN(59) TR(59)	0,13
	NN(68) TR(0)	0,22
	NN(68) TR(25)	0,21
	NN(68) TR(68)	0,1
	NN(111) TR(68)	0,18
	NN(111) TR(111)	0
	NN(331) TR(331)	0,13
EfficientNetV2-B0	NN(30) TR(30)	0,2
	NN(71) TR(30)	0,33
	NN(140) TR(140)	0,22
	NN(254) TR(254)	0,22
MobileNet	NN(35) TR(0)	0,19
	NN(54) TR(0)	0,2
	NN(72) TR(0)	0,26
ResNet50	NN(50) TR(38)	0,23
	NN(50) TR(50)	0,15
	NN(80) TR(38)	0,21
	NN(142) TR(142)	0,15
VGG-16	NN(10) TR(10)	0,18
	NN(14) TR(14)	0,22

False Color		
Network	Layer	Avg. d. diff.
EfficientNet-B2	NN(59) TR(0)	0,15
	NN(59) TR(25)	0,19
	NN(59) TR(59)	0,11
	NN(68) TR(0)	0,23
	NN(68) TR(25)	0,21
	NN(68) TR(68)	0,1
	NN(111) TR(68)	0,19
	NN(111) TR(111)	0
	NN(331) TR(331)	0,12
EfficientNetV2-B0	NN(30) TR(30)	0,21
	NN(71) TR(30)	0,31
	NN(140) TR(140)	0,22
	NN(254) TR(254)	0,2
MobileNet	NN(35) TR(0)	0,16
	NN(54) TR(0)	0,22
	NN(72) TR(0)	0,22
ResNet50	NN(50) TR(38)	0,22
	NN(50) TR(50)	0,15
	NN(80) TR(38)	0,23
	NN(142) TR(142)	0,12
VGG-16	NN(10) TR(10)	0,16
	NN(14) TR(14)	0,2

SWIR		
Network	Layer	Avg. d. diff.
EfficientNet-B2	NN(59) TR(0)	0,17
	NN(59) TR(25)	0,21
	NN(59) TR(59)	0,12
	NN(68) TR(0)	0,18
	NN(68) TR(25)	0,2
	NN(68) TR(68)	0,11
	NN(111) TR(68)	0,26
	NN(111) TR(111)	0
	NN(331) TR(331)	0,11
EfficientNetV2-B0	NN(30) TR(30)	0,21
	NN(71) TR(30)	0,3
	NN(140) TR(140)	0,21
	NN(254) TR(254)	0,19
MobileNet	NN(35) TR(0)	0,17
	NN(54) TR(0)	0,19
	NN(72) TR(0)	0,21
ResNet50	NN(50) TR(38)	0,2
	NN(50) TR(50)	0,15
	NN(80) TR(38)	0,17
	NN(142) TR(142)	0,14
VGG-16	NN(10) TR(10)	0,14
	NN(14) TR(14)	0,16

Different band composites of Central and Eastern Lithuania



True color image (TCI) from 2023

False color image from 2023

Short-wave infrared (SWIR) image from 2023

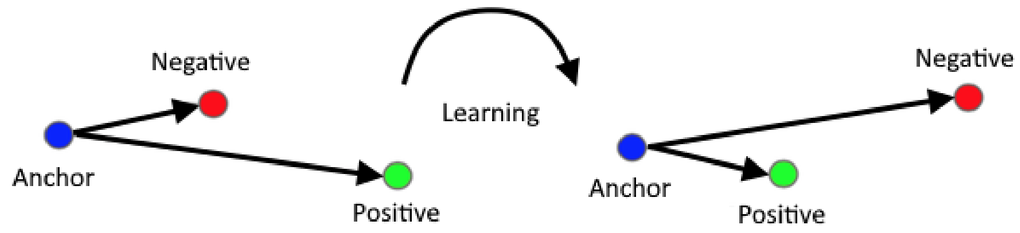
In order to create a natural colored output that accurately depicts the Earth as people would perceive it naturally, **true color** composite employs visible light bands red (B04), green (B03), and blue (B02) in the appropriate red, green, and blue color channels. The combination of the normal near infrared, red, and green band is used to produce **false color** images. False color composites with red, green, and near-infrared bands are quite common. Because plants absorb red light and reflect near-infrared and green light, it is most often used to evaluate the density and health of plants. Plant-covered terrain looks deep red because they reflect more near infrared than green light. Deeper red is the development of denser plants. Water looks blue or black, cities and exposed terrain are gray or brown. Water reflects **short-wave infrared (SWIR)** wavelengths, thus scientists may use these bands to determine the amount of water in plants and soil. The distinction between water and ice clouds, as well as snow and ice, which look white in visible light, may all be made using shortwave-infrared bands. SWIR bands reflect well from recently burnt ground, which makes them useful for mapping fire damage. Geology may be mapped by comparing reflected SWIR light because various types of rocks reflect SWIR light in different ways. In this composite, the blue channel displays the reflected red band, which highlights the built-up regions and bare soil, while the green channel displays B8A, which is reflected by vegetation.

Triplet Neural Networks

In some fields, such as reverse image search [3], human face recognition [4], and vehicle recognition from surveillance camera images [1], it is difficult but crucial to distinguish between similar and different image entities. It is useful to create a similarity measure that quantifies the shared information between given values [2] to estimate how similar two pictures are. Deep metric learning using Siam [12] and triplet [6] deep convolutional neural networks is one way to achieve this goal. Based on vector distance measurements, neural networks aim to learn from images and provide results that are close to similar images and distant from dissimilar images.

This is achieved in the case of Siamese networks by providing the network with two different images and a boolean indicating whether the images are members of the same class [8]. The network should increase the distances between different photos and decrease the distances between the feature maps of these photos when the boolean value indicates that the photos are similar.

A deep network is trained using triplet networks, which work in a similar way to Siamese neural networks [6], and tries to increase the distances between different pictures while decreasing the distances between similar pictures [9]. This is accomplished by providing the network with an anchor image (anchor) against which to compare two subsequent, positive images that are similar to the anchor or belong to the same or closely related class, and one negative image (*negative*) that is not similar to the main image or is in a class unrelated to the main image [6]. The network weights are modified so that the output distances between the main and positive images are smaller, and the distances between the main and negative images are larger [9] after computing the network outputs for all three images. This process is depicted in the figure below.



Triplet Loss

Results and conclusions

Five triplet neural networks with twenty-two different architectures were compared for True Color, False Color and SWIR band composites of Sentinel-2 imagery. Triplet neural networks perform similarly in terms of band composites individually, but the results vary between the different composites on Central and Eastern Lithuania area. Experiments were carried out with orthographic, satellite photos covering a large area, which define the influence of different bands combinations on the images' similarity detection in an area covering various areas, not only identified by the respective combination of bands. On average, true color composite showed the best results in terms of average distances (positive and negative distances) difference values, folowed up by false color composite. SWIR composite of Sentinel-2 imagery showed the least relevance on Central and Eastern Lithuania. Further reasearch on this topic will include areas like desert or arid regions, large bodies of water or snow-covered landscapes.

References

[1] Yan Bai et al. "Group-sensitive triplet embedding for vehicle reidentification". In: *IEEE Transactions on Multimedia* 20.9 (2018), pp. 2385–2399.

[2] Shihyen Chen, Bin Ma, and Kaizhong Zhang. "On the similarity metric and the distance metric". In: *Theoretical Computer Science* 410.24-25 (2009), pp. 2365–2376.

[3] Ye Chen. "Exploring the Impact of Similarity Model to Identify the Most Similar Image from a Large Image Database". In: *Journal of Physics: Conference Series*. Vol. 1693. IOP Publishing, 2020, p. 012139.

[4] Sumit Chopra, Raia Hadsell, and Yann LeCun. "Learning a similarity metric discriminatively, with application to face verification". In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 1. IEEE, 2005, pp. 539–546.

[5] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[6] Elad Hoffer and Nir Ailon. "Deep metric learning using triplet network". In: *Similarity-Based Pattern Recognition: Third International Workshop, SIMBAD 2015, Copenhagen, Denmark, October 12-14, 2015. Proceedings 3*. Springer, 2015, pp. 84–92.

[7] Andrew G. Howard et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications". In: *arXiv preprint arXiv:1704.04861* (2017).

[8] Iaroslav Melekhov, Juho Kannala, and Esa Rahtu. "Siamese network features for image matching". In: *2016 23rd international conference on pattern recognition (ICPR)*. IEEE, 2016, pp. 378–383.

[9] Florian Schroff, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 815–823.

[10] Mingxing Tan and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks". In: *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[11] Mingxing Tan and Quoc Le. "Efficientnetv2: Smaller models and faster training". In: *International conference on machine learning*. PMLR, 2021, pp. 10096–10106.

[12] Dong Yi et al. "Deep metric learning for person re-identification". In: *2014 22nd international conference on pattern recognition*. IEEE, 2014, pp. 34–39.