GENERATIVE ADVERSARIAL NETWORKS FOR HARD NEGATIVE MINING IN CNN-BASED SAR-OPTICAL IMAGE MATCHING

Lloyd H. Hughes[1], Michael Schmitt[1], Xiao Xiang Zhu[1,2]

[1]Signal Processing in Earth Observation, Technical University of Munich (TUM), Munich, Germany
[2]Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany

ABSTRACT

In this paper we propose a deep generative framework, based on a generative adversarial network (GAN) and an auto encoder (AE), for generating non-corresponding SAR patches to be used in hard negative mining in situations of limited data quantities. We evaluate the effectiveness of this formulation of hard negative mining for reducing the false positive rate (FPR) and improving network determinability in a SAR-optical patch matching application.

Our generative network is trained to generate realistic SAR images using an existing SAR-optical matching dataset. These generated images are then used as non-corresponding, hard negative samples for training a SAR-optical matching network. Our results show that we are able to generate realistic SAR images which exhibit many SAR-like features, such as layover and speckle. We further show that by fine tuning the original matching network using these hard negative samples we are able to improve the overall performance of the original SAR-optical matching network.

Index Terms— generative adversarial networks (GANs), synthetic aperture radar (SAR), multi-modal matching

1. INTRODUCTION

In recent years there has been a growing interest in utilizing deep learning methods to solve challenges in the field of remote sensing. With the increase in quality and availability of remote sensing data, these approaches have shown a lot of promise in tasks ranging from image segmentation through to speckle noise filtering [1].

One particularly interesting application which has benefited from the use of deep learning is that of finding correspondences between synthetic aperture radar (SAR) and optical remote sensing images of urban environments [2]. Apart from the photometrically challenging aspect of determining correspondence between two vastly different image manifolds, it also is required that matching is performed in such a way as to minimize the number of incorrect matches (false positive rate).

The discriminability of deep networks in classification tasks can be improved through the use of hard negative mining, a technique whereby false positives, or hard samples (samples which are statistically similar but have different classes) are used as negative examples during the learning phase [3]. In order for hard negative mining to be applied successfully, it is required that the training dataset is sufficiently large. This requirement is such that sufficient hard negative examples can be obtained when a classifier is performing with a high accuracy and low FPR. The lack of large scale SAR image datasets is thus a contributing factor to why hard negative mining has seen little application in deep learning for SAR imagery.

In this paper we investigate an alternative formulation of hard negative mining which does not rely on the existence of false positive samples, but rather on hard negative samples generated by generative adversarial networks (GANs). In our application to SAR-optical matching, for a given optical image, these hard negative samples are characterised as SAR images with a visual appearance similar to the true corresponding SAR image. These generated images, along with the corresponding SAR-optical pair, are then used to train a deep matching network. We theorize that using visually similar images as negative training samples will reduce the final false positive rate of the network, and will make the network more robust to real-world scenarios where multiple visually similar patches may occur in an image.

2. RELATED WORK

With the increase in use of deep networks for classification tasks in remote sensing, there has also been an increase in the application of hard mining to reduce false positive rates in these networks [4]. However, during our literature search it became clear that hard negative mining has only been applied to optical remote sensing classification tasks, where large scale datasets are more readily available. Furthermore, the use of GANs to generate hard negative samples for training was not found. The most similar application was the use of GANs to generate triplet pairs in a semi-supervised manner for use as training data in developing an image hashing network [5].

Furthermore, the use of GANs to generate SAR image data is still new area of research, although a few notable

examples already exist. Merkle et al. [6] utilized a well known conditional GAN architecture to generate synthetic SAR patches, of rural areas, from optical images in order to improve SAR-optical image matching. Another successful application was the use of GANs to simulate SAR images from digital 3D models of tanks [7] for use in automated target recognition. Finally, Marmanis et al. [8] used GANs to generate artificial SAR images in order to augment an existing dataset of annotated SAR imagery to improve SAR image classification. However, they failed to produce meaningful results when using these artificially generated images to augment their training data.

3. AUGMENTING SAR-OPTICAL MATCHING WITH GENERATED HARD SAMPLES

3.1. Matching SAR and Optical Patches with Pseudo-Siamese CNN

In order to evaluate our approach of using GANs to generate examples for hard negative mining we make use of the SAR-optical patch matching network proposed in [2]. This network has a pseudo-siamese architecture which takes a SAR patch and an optical patch as input and outputs a binary value indicating correspondence. The network was trained on a dataset of corresponding TerraSAR-X and UltraCam image patches pairs[9][2].

This network was shown to have a similar performance to standard signal based methods, with an accuracy of 77% being achieved at a false positive rate (FPR) of 5%. The FPR was set to a maximum of 5% in order to ensure robustness in real-world applications.

In order to assess the success of hard negative mining in improving the false positive rate, the original network was fine tuned using the same hyper-parameters and training dataset, except with a modification to the non-corresponding pairs. Instead of using the originally randomly assigned non-corresponding SAR patches, we replaced these with generated patches which are visually similar to the original corresponding SAR patches. An example of the training set we used can be seen in Fig. 1.

3.2. Generating Similar SAR images

In order to generate SAR images for use as hard negative samples, it is required that we can generate realistic SAR images in a controlled manner, such that they are visually similar to the original image. Unlike other applications in remote sensing which have generated SAR images in a conditional manner [6][7], or have required entirely unique images [8], our application has the additional requirement that generated images need to be both novel and structurally similar to the original image.

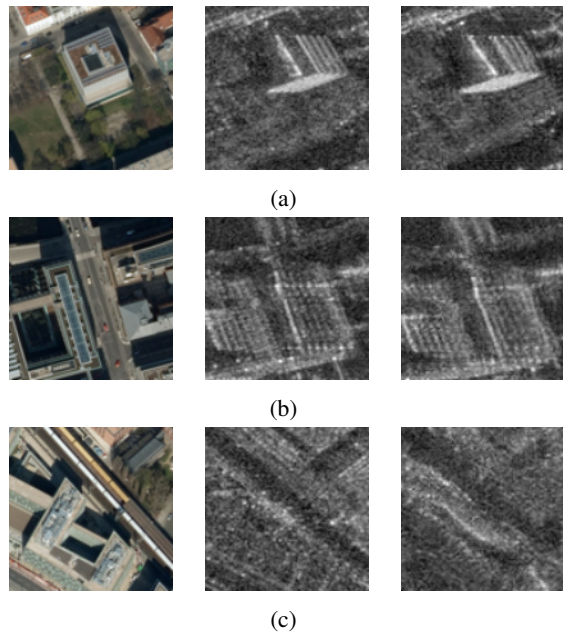In order to achieve this we modified the deep convolutional GAN (DCGAN) architecture, proposed in [10], to

Fig. 1: Example images from our training dataset. The left column shows the true optical image and the middle column the corresponding SAR image, the right column depicts the generated hard negative sample.

include an additional encoder network E. Coupled with the generator G, the encoder network is trained as an auto-encoder, to reconstruct the original SAR image patch. This architecture ensures that a latent code z exists which describes the original SAR image and ensures some degree of smoothness in the latent space. The full generative architecture can be seen in Fig. 2, while the inference network can is depicted in Fig. 3.
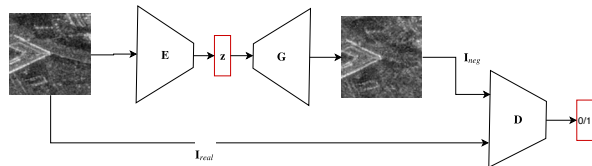

Fig. 2: Our modified DCGAN architecture. Network E and G form an auto-encoder which helps define the latent space z, while networks G and D are trained in a standard GAN manner.

The training of our generative network was performed in two stages. Firstly the auto-encoder networks E and G are trained using an $L2$ reconstruction loss. The GAN networks, G and D, are then trained using the input vector z created by combining the latent code generated by the encoder network with a Gaussian noise vector. This modified GAN formulation is then trained in an end-to-end manner using the Wasserstein GAN gradient penalty loss (WGAN-GP) proposed by

[11]. The WGAN-GP loss has been shown to improve training stability and reduce the chance of mode collapse.
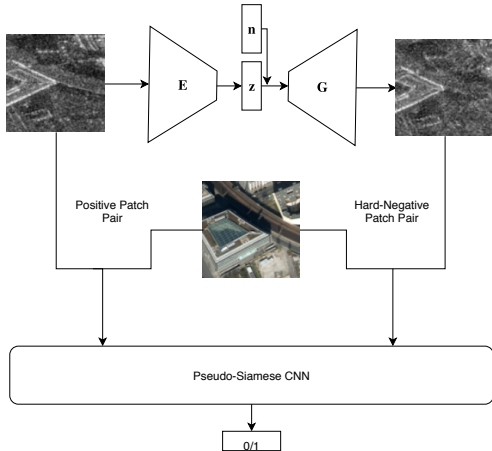

Fig. 3: When performing inference we only make use of the E and G networks in order to generate hard negative samples. These samples are then used to train the matching network.

3.2.1. Progressive Growing of GANs

Training GANs can be a challenging tasks due to the inherent instability in the adversarial learning task. This instability is exacerbated by the large amounts of speckle noise and complex imaging geometry found in SAR data. In previous attempts to train GANs for generating SAR images, all the authors mentioned the difficulties they faced in training the various GANs due to training instability [6][7][8]. Furthermore, Marmanis *et al.* [8] mentioned the difficulty in generating large SAR images which were required for their application.

In order to improve the GAN's stability during training, we make use of a progressive growing technique proposed by [10]. In this approach the resolution of generator and discriminator networks is progressively increased at fixed intervals during training. The generator starts off by generating a low resolution sample from a random latent code, which is passed into a low resolution discriminator network. This discriminator network attempts to recover the original latent code. After a fixed duration an additional layer is added to both networks, thereby doubling the resolution of the generated image. We further extended this progressive growing formulation to include growing the encoder network resolution in conjunction with the GAN network. An example of the GAN's training progression can be seen in Fig. 4.

3.2.2. Generating Hard Examples

To generate hard negative training samples, we exploit the continuous latent space learned by our AE-GAN framework. A hard negative training sample is created by encoding the
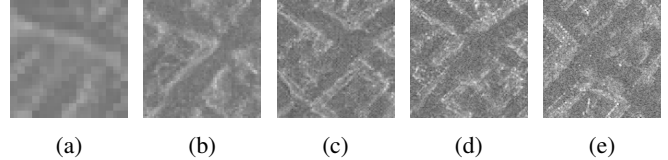

Fig. 4: An example of the progressively grown images taken at increasing image resolutions, (a)-(d): generated images with increasing resolution from 16×16 pixels up to 128×128 pixels; (e) A real, 256×256, SAR image patch form the training set.

original SAR image, $\mathbf{I}_{SAR}$, to obtain its latent representation $\mathbf{z}$. A small amount of noise $\mathbf{n}$ is then added to $\mathbf{z}$ to create a unique latent representation near to the original code. This is then used by G to generate an image with a similar distribution but slightly different structure, which is then used as a hard negative training sample. The process of generating image pairs for training the pseudo-siamese matching network [2] is depicted in Fig. 3.

4. EXPERIMENTS AND RESULTS

4.1. Generated Hard Negative Pairs

Examples of hard-negative samples generated by our AE-GAN compared to the original patch pair are depicted in Fig. 1. It is clear from these examples that our network has successfully learnt how to generate realistic SAR images, which exhibit natural features of SAR imagery such as speckle, layover, and shadow. Furthermore, on closer inspection of these generated images it is clear that they are structurally similar to the original SAR patches. The generated image presented in Fig. 1c depicts a failure case of our network where the generated patch is obviously different from the original patch and does not constitute a hard-negative sample.

4.2. Effects on SAR-optical Matching

from these examples In order to evaluate the effects of our generated hard negative samples on the false positive rate and overall accuracy of the SAR-optical matching network described in [2], we fine-tuned the original network using our modified dataset described in Sec. 3.1. The results of this fine tuned network are presented in Tab. 1, where they are compared with the results of the original network trained using the original dataset.

5. DISCUSSION

From the results presented in Tab. 1 it can be seen that the inclusion of hard negative examples provides the matching network with a significant improvement in terms of its ability

Table 1: Results describing the influence of training with hard negative examples on accuracy and false positive rates (FPR).

Experiment	Accuracy	Precision	Recall	FPR
Original with random neg.	82%	0.87	0.75	11.7%
Fine-tuning with hard neg.	78%	0.93	0.61	4.0%

to distinguish negative samples, significantly decreasing the false positive rate.

Interestingly, the overall matching accuracy decreases slightly. This can probably be explained by a reduction of overfitting caused by a lack of heterogeneity in the original SAR-optical patch-pair dataset. Since the accuracy in the original paper was about 77% when restricting the FPR to a maximum of 5% [2], our hard negative mining extension actually results in a slight improvement to matching accuracy at significant reduction of the FPR. Thus, overall the fine-tuned network exhibits behaviours which are preferable for matching applications where outliers can have detrimental effects on data fusion products which rely heavily on the quality of the matching results.

6. SUMMARY AND CONCLUSION

In this paper we have shown an approach for reducing the FPR of a matching application when insufficient quantities of data are available for standard hard negative mining techniques. Using a deep AE-GAN architecture we were able to generate realistic SAR image patches, which are both novel and structurally similar to the original SAR patches in our training dataset. These generated images were shown to work well as hard negative examples for fine tuning an existing matching network in order to achieve better determinability and a lower FPR in a matching task.

Future work will aim at improving the interpretablity of the latent codes produced by the auto encoder network to enable better controllability during the creation of novel images. Furthermore, we will investigate the possibilities of training the AE-GAN and matching network in an end-to-end manner using a triplet training loss.

Acknowledgements

References

[1] X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A review," *IEEE Geosci. Remote Sens. Mag.*, 2017.

[2] L. H. Hughes, M. Schmitt, L. Mou, Y. Wang, and X. X. Zhu, "Identifying corresponding patches in sar and optical images with a pseudo-siamese cnn," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 784–788, 2018.

[3] K.-K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 1, pp. 39–51, 1998.

[4] T. Tang, S. Zhou, Z. Deng, H. Zou, and L. Lei, "Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining," *Sensors*, vol. 17, no. 2, p. 336, 2017.

[5] Z. Qiu, Y. Pan, T. Yao, and T. Mei, "Deep semantic hashing with generative adversarial networks," in *Proc. ACM SIGIR*, Tokyo, Japan, 2017, pp. 225–234.

[6] N. Merkle, P. Fischer, S. Auer, and R. Mller, "On the possibility of conditional adversarial networks for multi-sensor image matching," in *Proc. IGARSS*, Fort Worth, USA, 2017, pp. 2633–2636.

[7] J. Guo, B. Lei, C. Ding, and Y. Zhang, "Synthetic aperture radar image synthesis by using generative adversarial nets," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1111–1115, 2017.

[8] D. Marmanis, W. Yao, F. Adam, M. Datcu, P. Reinartz, K. Schindler, J. D. Wegner, and U. Stilla, "Artificial generation of big data for improving image classification: A generative adversarial network approach on SAR data," in *Proc. BiDS*, Toulouse, France, 2017, pp. 293–296.

[9] Y. Wang, X. Zhu, B. Zeisl, and M. Pollefeys, "Fusing meter-resolution 4-D InSAR point clouds and optical images for semantic urban infrastructure monitoring," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 1–13, 2016.

[10] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. ICLR*, Vancouver, Canada, 2018.

[11] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. NIPS*, Long Beach, USA, 2017, pp. 5769–5779.