Research Statement
of Linlin Yang to TUM computer vision group

I am broadly interested in computer vision with particular emphasis on image reconstruction and real-time image processing. My past research, aiming to better understand the inherent data structure, can be broadly classified into two areas, structure constraint and deep learning. In future, I would like to broaden my research field and bridge the gap between 2D and 3D. In this statement, I briefly summarize the contributions of my prior research, and detail my future research plans.

Structure Constraint. Imposing a data structure as a constraint is a new and flexible way to solve the optimization problems, and it has shown a lot of promise in improving performance of machine learning algorithms. This constraint brings robustness to the variations and have been explored in many field.

The first time I learned this method was from my senior's work. In order to devise a stable long-term tracking scheme, we focused on the data distribution. According to the observation, we found that the correlation response of the target image lay on a Gaussian distribution and the noisy samples, which caused drifting, might degrade this distribution. With this observation, we suggested that the distribution might enhance the robustness and monitor the drift. So we imposed a data structure (Gaussian distribution) as a prior constraint into the kernelized correlation filter (KCF) and achieved an excellent performance. We finished a rough version [1], and then my senior well investigated this work in both theory and experiment [2].

Encouraged by this success, I started to learn more about structure constraint and attempted to apply it to margin theory. In contrast to the conventional max-margin method, I thought that the distribution could be used as a sample selection strategy. After discussion in our group, my advisor suggested me to investigate this method in ensemble learning because of its potential to reduce over-fitting. With my advisor's help, we did some experiments and proved that Gaussian mixture model (GMM) distribution constraint could maximize margins and meanwhile control the margin variance. We applied this method to human action recognition, and the proposed system was superior to the state-of-the-art. This paper was accepted to IEEE TIP [3].

Even though structure constraint is popular in detection, tracking and classification, it's rare to use in the image restoration or reconstruction. I think this constraint has great potential to reduce the perturbation of the input so that it could improve the performance of the restoration or reconstruction. So I focused on this direction and aimed to solve video or batch image problems like the light field reconstruction and the misalignment problem. I studied one of my professor's work, manifold constraints transfer (MCT) [4] and proposed a similar idea. We knew that the data often lay on a specific and unknown manifold. So we assumed the data lay on such a manifold and then added a manifold constraint as an unsolved part to the original problem. With the framework of alternating direction method of multipliers (ADMM), we solved the new problem iteratively and ended up with a steady structure based on its "true" neighbors so to avoid perturbations from samples far from the input data. We used linear local embedding (LLE) and MCT to handle misalignment problems and achieved much better results than original methods. This paper was submitted to IEEE TCSVT [5]. Differently, I also proposed another coherent framework to process the reconstruction problem and achieved a robust result. This work was accepted to IEEE JSTSP [6].

Deep Learning. I have paid much attention on deep learning. Personally speaking, I think deep learning is to extend conventional algorithm to hierarchical structure to handle more complicated information. There are many details like the structures, the filters or the optimization algorithms need to be explored. My group have done some amazing works and I also have learned two of them.

The first one is about the filters. When we did some experiment on aerial photography dataset, we found deep convolutional neural networks (DCNNs) lack the capability to handle large image rotation. Inspired by SIFT and Gabor features, our group proposed two methods to solve this problem. Zhou *et al.*[7] proposed active rotating filters to extend the original convolutional filters to multi-channel filters with rotation information. With the SIFT-like filters, DCNNs could be applied to process the rotation-invariant data and estimate the object orientation. Meanwhile, Luan *et al.*[8] proposed Gabor orientation filters to improve the robustness of DCNNs to image transformations. The two works both led to significant reduction in the number of network parameters and improvement in classification performance.

Another one is about the structure constraint, which is an effective way to improve the performance of DCNNs. Li *et al.*[9] introduced spatio-temporal manifold as a regularization term into the loss function. It transferred the structure of the data to the variable in DCNNs and showed significant improvement to the baselines. Differently, Su *et al.*[10] designed a new architecture with singular value decomposition(SVD) layer to satisfy the intrinsic manifold structure and also achieved the state-of-the-art results.

Future Plan. In future, I would like to continue investigating structure constraint and deep learning, and explore some new fields like 3D reconstruction.

First, I plan to study how to apply structure constraint to video application, especially video retrieve or detection. Additionally, I think structure constraint now particularly depends on the experience or observation, which limits its potential. We should introduce structure constraint to metric learning, transfer learning or deep learning. For example, when it comes to generative adversarial networks (GAN), I think that structure, the consistence of the data, may be the connection between discriminator and generator. As a prior constraint, it will be a powerful regularization to reduce over-fitting or perturbation.

Another probable direction of my future work is the output layer of deep learning. Muti-channel correlation filters can be seen as a one-layer deep learning with Gaussian output. So I think it's a wonderful idea to develop deep correlation filter or we may call it deep learning with Gaussian output. Moreover, I think output layer also have very large development potential, like the convolution operator and special-purpose layers of deep learning. For example, we can add Gaussian output or structured output [11] to DCNNs to process video tasks. In my opinion, a flexible output layer develops DCNNs to tackle some more complicated situations.

Last but not least, I want to mention convolutional sparse coding (deconvolution) which is superior to the traditional patch based sparse coding. It's state-of-the-art method to 2D or 3D reconstruction which already has applied to GAN. According to my experience, 3D deconvolution for 3D reconstruction is worth studying. Moreover, because of its interpretability [12][13], it also can be a complementary method for deep learning.

Conclusion. Thanks to my advisor Pro. Baochang Zhang and co-advisor Pro. Qixiang Ye, I have learned how to conduct a research independently. I believe I will refine my thoughts and continue exploring new areas in the following study and research. TUM computer vision group is an ideal lab with great works and I hope to have the opportunity to join in.

References

[1] X. Su, B. Zhang, L. Yang, Z. Li, and Y. Yang, "Scale invariant kernelized correlation filter based on gaussian output," in *International Conference on Cloud Computing and Security*. Springer, 2016, pp. 567–577.

[2] B. Zhang, Z. Li, X. Cao, Q. Ye, C. Chen, L. Shen, A. Perina, and R. Jill, "Output constraint transfer for kernelized correlation filter in tracking," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 4, pp. 693–703, 2017.

[3] B. Zhang, Y. Yang, C. Chen, L. Yang, J. Han, and L. Shao, "Action recognition using 3d histograms of texture and a multi-class boosting classifier," *IEEE Transactions on Image Processing*, 2017.

[4] B. Zhang, A. Perina, V. Murino, and A. Del Bue, "Sparse representation classification with manifold constraints transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4557–4565.

[5] B. Zhang, L. Yang, J. Han, V. Murino, L. Jianzhuang, and A. Del Bue, "Manifold constraint for rotation invariant low rank decomposition," *IEEE Transactions on Circuits and Systems for Video Technology*.

[6] L. Yang, C. Li, J. Han, Q. Ye, C. Chen, X. Cao, B. Zhang, and W. Liu, "Manifold constrained convolutional sparse coding for image sets," *IEEE Journal of Selected Topics in Signal Processing*, 2017.

[7] Y. Zhou, Q. Ye, Q. Qiu, and J. Jiao, "Oriented response networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[8] S. Luan, B. Zhang, C. Chen, X. Cao, Q. Ye, J. Han, and J. Liu, "Gabor convolutional networks," *arXiv:1705.01450*, 2017.

[9] C. Li, C. Chen, B. Zhang, Q. Ye, J. Han, and R. Ji, "Deep spatio-temporal manifold network for action recognition," *arXiv:1705.03148*, 2017.

[10] X. Su and B. Zhang, "Deep structure transfer machine theory and applications."

[11] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun, "Large margin methods for structured and interdependent output variables," *Journal of Machine Learning Research*, pp. 1453–1484, 2005.

[12] V. Papyan, J. Sulam, and M. Elad, "Working locally thinking globally: Theoretical guarantees for convolutional sparse coding," *arXiv:1707.06066*, 2017.

[13] P. Vardan, Y. Romano, and M. Elad, "Convolutional neural networks analyzed via convolutional sparse coding," *arXiv:1607.08194*, 2016.