

# Critical Review Of Seminars

**Author:** Yao Xiao

**Supervisor:** Xinguo Yu

**Student Number:** 2019180015

Because of my strong interest in computer vision, I chose two lectures from Professor L.Wang of the University of Wollongong, and conducted relevant research and reflection on the content of the two lectures.

## Review On Research Topics In The VILA Group Of SCIT In UOW

In the lecture, Professor Wang shared the recent research directions, such as comics recognition and image retrieval. For CRM-based image retrieval [2], Professor Wang proposed a method to solve this problem, which regards the locally clustered image regions as pseudo-labeled classes that share the same visual concept, and used them to model the distribution of visual concept classes associated with image regions. Both non-parametric methods and parametric methods have been developed for this purpose and have been carefully demonstrated by probabilities. Extensive experimental studies on multiple benchmark datasets have shown that the proposed pseudo-label method performs well in CRM and other comparable methods, compared with CRM, the biggest improvement is over 10 percentage points. Professor Wang also mentioned that due to the heterogeneity of increasingly complex scientific data, new complexity methods were needed. Interactive visual analysis was still a new interdisciplinary field that would analyze and interactive visual methods. Schemes such as interactive linked provide powerful insight into mechanisms to represent the information represented.

In fact, in my views, visualization and visual analysis play an important role in exploring, analyzing, and presenting scientific data. In many disciplines, data and model approaches are becoming multifaceted: data is often spatiotemporal and multivariate. They come from different data sources, multiple multiphysics simulations of simulation runs or interaction phenomena. Similarly, the data can be constructed on different types of grids or on various types of meshes that need to be associated or fused in the visualization. This heterogeneity of data characteristics brings new opportunities and technical challenges to visualization research. Therefore, visualization and interaction techniques are often combined with computational analysis.

In this case, an important step is to integrate user input and data from machine learning methods that can be learned from previous courses, and configure parameters for visualization based on the acquired knowledge.

But at the same time, there are still many problems, such as extracting meaningful information from heterogeneous scientific data. Such data can be fused at the data level, for example by resampling or by using a different model of the data model that provides uniform access. Another option is to fuse multi-faceted data at the functional level, for example, by exchanging selection information across different data sections.

Although current methods rarely address the heterogeneity of multifaceted scientific data. But here we see that absolutely new concepts and methods are absolutely needed, so it can be regarded as a promising research direction in visualization.

## **Review On Learning Second-Order Feature Representation For Visual Recognition**

Professor Wang's speech suggested that recently learning high-level feature representations has attracted a lot of attention in computer vision. As a second-order merged representation, the covariance matrix played an important role in this research trend. In this case, the kernel matrix is even better than the covariance matrix by modeling higher-order non-linear relationships between the combined visual descriptors based on previous research. In addition, the talk reported on recent work on developing a deep learning network that will jointly learn local descriptors and pooled representations based on kernel matrices in an end-to-end manner. The experiments performed on the fine-grained image benchmark dataset show that the SPD representation based on the kernel matrix has deeper local descriptors and is superior, and validates the advantages of the proposed deep web word in pursuit of a better SPD representation.

Actually, to handle image changes, modern visual recognition typically models the appearance of an image using a set of local descriptors. They evolved from early filterbank responses (through traditional local invariant features) to the activation feature maps of recent deep convolutional neural networks (CNNs). In this process, how to gather a set of local descriptors to obtain a global image representation has become a central issue, and many excellent methods have been proposed in the literature. In the past few years, using a covariance matrix to merge a set of descriptors has attracted more and more attention, and has shown encouraging results in terms of object recognition, image set classification, etc. It describes the pairwise correlation of descriptor components, and because the covariance matrix is SPD, it is often called a symmetric positive definite (SPD) representation. Moreover, this inspired research on classification, clustering and dimensionality reduction of SPD representations. In particular, some recent pioneering work has integrated this SPD representation based on the covariance matrix into deep CNNs to learn the covariance matrix together with local visual descriptors. These works have studied several important issues in this deep learning framework, including deriving some matrix-

based back-propagation functions, the correct method of normalizing the covariance matrix, and helping large-scale visual recognition of second-order information.

As far as I can tell, as a second-order merged representation, the covariance matrix has attracted wide attention in visual recognition, and some pioneering work has recently integrated it into deep learning. Many recent studies have shown that by modeling higher-order non-linear relationships between merged visual descriptors, the kernel matrix works much better for this representation than the covariance matrix. However, in this study, neither the descriptors nor the kernel matrix were deeply understood. Therefore, how to show the superiority of SPD representation and deep local descriptor based on kernel matrix in subsequent research, and how to verify the advantages of the proposed deep network in pursuit of better SPD representation. In addition, how to provide ablation studies to explain why and where to obtain these improvements is the focus and direction of future research.

## References

- [1] Wang, Yan, Luping Zhou, Biting Yu, Lei Wang, Chen Zu, David S. Lalush, Weili Lin, Xi Wu, Jiliu Zhou, and Dinggang Shen. "3D auto-context-based locality adaptive multi-modality GANs for PET synthesis." *IEEE transactions on medical imaging* 38, no. 6 (2018): 1328-1339.
- [2] Li, Wenbin, Jing Huo, Yinghuan Shi, Yang Gao, Lei Wang, and Jiebo Luo. "A Joint Local and Global Deep Metric Learning Method for Caricature Recognition." In *Asian Conference on Computer Vision*, pp. 240-256. Springer, Cham, 2018.
- [3] Gao, Zhimin, Lei Wang, and Luping Zhou. "A Probabilistic Approach to Cross-Region Matching-Based Image Retrieval." *IEEE Transactions on Image Processing* 28, no. 3 (2018): 1191-1204.