A Strategical Analysis and A Comparative Study on DMF

*K.Elaiyaraja[1], Dr.M.Senthil Kumar[2], Dr.B.Chidhambararajan[3]
[1]*Department of IT,* [2]*Department of CSE,* [3]*Principal/Professor*
[1,2,3]*SRM Valliammai Engineering College*

Abstract: *Nowadays deep learning plays an important role in various understanding tasks. Generally, multiple-modalities offerharmonizing data on the identical scene. A variety of studies shows that deep-multimodal fusion achieves important or notable performance development. These fusion techniques have the advantages of variety of data sources and can generate optimized predictions robotically. This work elaborates the important background technical concepts of DMF (Deep Multimodal Fusion) and existing applications used which is related with it. In addition to that, a systematic assessment on multimodal fusion techniques are provided. A comparative study on existing fusion techniques, its strength and weakness along with its performance and challenges are analyzed.*

Keywords: Fusion, Image Fusion, Segmentation, Deep learning, Multi modal fusion

1. Introduction

In computer era, semantic segmentation is considered as a higher grade task. Image segmentation leads to assigning a semantic tag for every pixel exist in the source images. Instance level segmentation generates mask tag and class tag for each instance. Panoptic segmentation technique is a familiar approach which integrates instance as well as pixel level segmentation*[11]*. There are variety of machine learning methods are proposed to handle these challenges with the help of DL methodologies. CNN (Convolutional Neural Networks) is a popular technique for pattern recognition and RNN (Recurrent Neural Networks) retrieves contextual information. The evolution of DL sets a new milestone in this image segmentation.

DMF techniques benefited from the huge amount of data and significantly increases the computation. These fusion techniques provide accurate results with redundant scene data. Extracting significant information by designing compact technologies to enhance the targeted data*[12]*. DMS (Deep Multimodal Segmentation) is to acquire the significant features of the identical scenes. It is necessary to improve the robustness and accuracy of DMF and scalability as well as the challenges faced in real times are considered for the real time based applications.

Image fusion obtained from multiple sources and acquires significant data from the identical scenes. The main agenda of multimodal fusion is to attain significant information by integrating identical scenes from different modalities. In medical era, Positron Emission Tomography (PET), Computed Tomography (CT), Single-Photon Emission Computed Tomography (SPECT), etcetera is used for multi modal image fusion*[14]*. Diagnosing accurate disease is possible by using DMF in medical area where as in remote sensing oriented applications, biometric identification and vehicle interactions are tackled.

2. Semantic Image Segmentation

The n number of studies provides the process of image segmentation using DL methodologies. FCN (Fully Convolutional Network) was proposed initially for dealing with pixel level classifications. Here, the fully connected level is replaced with convolutional layers. In DeconvNet*[10]*, the deconvolution and un-pooling substances are dealt. An encoder and decoder

framework with forward pooling was introduced in *[1]* and stated as SegNet. U-Net is an another segmentation network proposed to handle biomedical related image segmentation. It consists of combining semantic information which is retrieved from decoder with least level adequate grained data of the encoder. The ENet concept was proposed for real-time segmentation. The Bansal et al *[2]*, proposed PixelNet which explores the spatial coordination between pixels in order to improve the performance and efficiency of segmentation methodologies. The DilatedNet was presented to retain the feature resolution which maps with receptive areas. The Deeplab techniques also attains excellent targeted information on image segmentation.

3. Deep Multimodal Segmentation

In earlier days, integrating information obtained from various data sources into low dimensional area was done as data fusion or information fusion. Including ICA (Independent-Component Analysis) and CCA (Canonical-Correlation Analysis) were used as fusion techniques in Machine Learning. The discriminative classifierwas a popular technology in combining multimodal data and it is called as decision or late fusion. Before CNN, these fusion techniques were convincing techniques for a long time.

DL techniques have reasonable advantages in performance and learning ability when compared with Machine Learning. Roughly, DMF techniques enhance the unimodal concept with an optimized fusion strategy*[16]*. These unimodal concepts are derived from the traditional methodologies which represents UNN (Unimodal Neural Networks) like VGGNet and ResNet. In DMF, the semantic approach of image segmentation is considered as initial step which trains the concatenated data on a SNN (Single Neural Network).

4. Understanding Postures

The main task to be faced in DMF is impostures issues. Hazirbas et al *[6]* presents the issues of pixel level predictions with the help of color and depth information. Schneider et al *[13]* introduces a framework called mid-level fusion for metropolitan image segmentation. Here, indoor and outdoor image segmentations are used. Valada et a*[15]* proposed posture understanding of formless environments like forests. There are so many posture understanding scenarios got benefited from multimodal fusion like object tracking, humanoid detection, trip-hazard detection, Salient-object detection. LiDAR is always used to provide highly optimized 3D cloud data. Petel et al *[11]* addresses fusion using RGB and three dimensional LiDAR information for indoor related circumstances. Recently so many techniques were proposed for adopting cloud maps and mainly focusing on three dimensional based object identification. DMF has identical and mixed data sources and they can be a robust prominence for intellectual mobility in future.

5. Fusion Approaches

5.1Classification

Generally, classification of fusion including data, early, late, intermediate and hybrid fusion follows various strategies. In early fusion technique, it focuses on concatenation of raw data from various modalities into different channels. The learning methodology is trained from end to end with a single segmentation system. The segmentation system adapts the information via equivalent section followed by an additional unit is hired to calculate the weights and the encoder and decoder acquaintances.

5.2 Early Fusion

Couprie et al *[3]* invented a DMF in the year 2013. It elaborates the early fusion with RGB as well as depth channels before segmentation system process. This process providespromising results for indoor posture recognition in the case of identical depth presence and site. But it has limitations of

multi modal data extraction for a simple concatenated images. Hu et al *[7]* presented an early fusion technique by ACNet which gathers significant features from RGB as well as depth sectors. RTFNet was specially designed for integrating RGB and thermal kinds of images with the help of element-wise abstraction. In order to boycott huge loss of spatial data, the average pooling and complete connected layers were employed in the system.

5.3 Late Fusion

Gupta et al *[5]* invented a mechanism for object recognition and segmentation. Two CNN streams were used to extract RGB and depth information. The SVM classifier was used to combine and obtain feature maps from these. The LSTM-CF (Long-Short Term Memorized Context-Fusion) model was developed by Li et al *[9]* for semantic tag of RGB-D postures. This system obtains photometric and depth data parallel along with facilitation of deep fusion of contextual data. The universal contexts and the least convolutional data of RGB stream are integrated by simple concatenation. More lately, CMnet*[17]* done to discover the corresponding features of polar metric information. Here, various backbone systems were used to extract multimodal information. The following Figure 1 depicts the performance of various DMF techniques.

Techniques	Input size	Backbone	Mean Acc	Fusion Technique	Mean IoU
SSMA	768*384	ResNet-50	-	Hybrid	44.52
RedNet	640*480	ResNet-50	60.3	Hybrid	47.8
CFN	-	RefineNet-15	-	Hybrid	48.1
FuseNet	224*224	VGG-16	48.3	Early	37.3
RDFNet	-	ResNet	60.1	Early	47.7
LSTM-CF	426*426	VGG-16	48.1	Late	-
Context	-	VGG-16	53.4	-	42.3
LSD-GF	417*417	VGG-16	58	Late	-
RefineNet	-	ResNet-152	58.5	-	45.9
DFCN-DCRF	480*480	VGG-16	50.6	Early	39.3

Figure 1: The performance of popular DMF techniques

5.4 Hybrid fusion

According to the previous studies, integrating weighted features is not sufficient at decision level to fulfil the accurate result. The hybrid strategy is invented to fuse the strength of legacy and late fusion. Residual learning was the core concept in earlier fusion methods and it is called as RDFNet which combines RGB-D features*[4]*. This one leads to high-definition prediction by the feature fusion and refinement class. The RGB part produces feature mappings used to map the regions which complements sectors. The CaRF (Context-Aware Receptive Field) empowers the fusion system to get a competitive block output. In addition to that, S-M fusion was introduced to study the feature representation through bottom-up approach.

5.5 Statistical fusion

In order to reduce the uncertainty at decision level based fusion, statistical fusion was initiated. Statistical fusion combines deep learning segmentation prediction, Dirichlet fusion and Categorical fusion*[8]*. Deprived of extra training data, a small subset is required for calibrating the statistical system. Merging more classifiers in this fusion is not a latest technique but this approach gives an exciting research way to merge statistics and deep learning.

6. Discussion

DMF for understanding posture issues is a tedious process which includes spatial positioning of an object, the semantic context of image, fusion system effectiveness, physical properties etc. The above specifies fusion strategies follows various design techniques to handle this kind of challenges. Legacy fusion techniques optimally combines data retrieved from source images. Late fusion techniques usually correlate multimodal features into a communal space. That is unimodal features trains the fusion system in a separate manner. Flexibility and scalability can be achieved in late fusion with limited cross model correlation. In hybrid fusion, such kind of fusion techniques integrate the effectiveness of early and late fusion which gives more robustness.

7. Conclusion

So far, DMF was reviewed for semantic image segmentation which elaborates the enhancement of multimodal fusion and makes the reader to learn with background acquaintances. Then, DMF categories are provided as early, late and hybrid fusion and additionally covered the importance. The DMF gains more attention recently in last decades. The experimental output was collected in this review and shows the efficiency of DMF techniques. But the optimal fusion yet to explore further. DMF based AI is steadily evolving from perception to cognitive approach and the DMF simplifies this development. In the upcoming years, this evolution offers a mass innovation.

REFERENCES

[1] *Badrinarayanan V, A. Kendall, R. Cipolla, SegNet: A Deep Convolutional Encoder Decoder Architecture for Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2017) 2481–2495.*

[2] *Bansal, X. Chen, B. Russell, A. Gupta, D. Ramanan, Pixelnet: Towards a general pixel-level architecture, arXiv preprint arXiv:1609.06694 (2016).*

[3] *Couprie C, C. Farabet, L. Najman, Y. LeCun, Indoor Semantic Segmentation using depth information, arXiv preprint arXiv:1301.3572 (2013).*

[4] *Deng L, M. Yang, T. Li, Y. He, C. Wang, RFBNet: Deep Multimodal Networks with Residual Fusion Blocks for RGB-D Semantic Segmentation, arXiv preprint arXiv:1907.00135 (2019).*

[5] *Gupta S, R. Girshick, P. Arbeláez, J. Malik, Learning rich features from rgb-d images for object detection and segmentation, in: European conference on computer vision, Springer, 2014, pp. 345–360.*

[6] *Hazirbas C, L. Ma, C. Domokos, D. Cremers, FuseNet: Incorporating depth into semantic segmentation via fusion-based CNN architecture, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 10111 LNCS, 2017. doi: 10.1007/978-3-319-54181-5_14.*

[7] *Hu X, K. Yang, L. Fei, K. Wang, ACNet: Attention Based Network to Exploit Complementary Features for RGBD Semantic Segmentation, arXiv preprint arXiv:1905.10089 (2019).*

[8] *Li C, D. Song, R. Tong, M. Tang, Illumination-aware faster r-cnn for robust multispectral pedestrian detection, Pattern Recognition 85 (2019) 161–171.*

[9] *Li Z, Y. Gan, X. Liang, Y. Yu, H. Cheng, L. Lin, LSTM-CF: Unifying context modeling and fusion with LSTMs for RGB-D scene labeling, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 9906 LNCS, Springer, 2016, pp. 541–557. doi: 10.1007/978-3-319-46475-6_34. arXiv:1604.05000.*

[10] *Noh H, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, in: Proceedings of the IEEE International Conference on Computer Vision, volume 2015 Inter, 2015, pp. 1520–1528. doi:10.1109/ICCV.2015.178. arXiv:1505.04366.*

[11] *Patel N, A. Choromanska, P. Krishnamurthy, F. Khorrami, Sensor modality fusion with CNNs for UGV autonomous driving in indoor environments, in: IEEE International Conference on Intelligent Robots and Systems, volume 2017-Septe, IEEE, 2017, pp. 1531–1536. doi: 10.1109/IROS.2017.8205958.*

[12] *Qiu S, Q. Fu, C. Wang, W. Heidrich, Polarization demosaicking for monochrome and color polarization focal plane arrays (2019).*

[13] *Schneider L, M. Jasch, B. Fröhlich, T. Weber, U. Franke, M. Pollefeys, M. Rätsch, Multimodal neural networks: RGB-D for semantic segmentation and object detection, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 10269 LNCS, Springer, 2017, pp. 98–109. doi: 10.1007/978-3-319-59126-1_9.*

[14] *Taghanaki S A, K. Abhishek, J. P. Cohen, J. Cohen-Adad, G. Hamarneh, Deep semantic segmentation of natural and medical images: A review, arXiv preprint arXiv:1910.07655 (2019).*

[15] *Valada A, G. L. Oliveira, T. Brox, W. Burgard, Deep Multispectral Semantic Scene Understanding of Forested Environments Using Multimodal Fusion, in: International Symposium on Experimental Robotics, Springer, 2017, pp. 465–477. doi: 10.1007/978-3-319-50115-4_41.*

[16] *Xiao Y, F. Codevilla, A. Gurram, O. Urfalioglu, A. M. López, Multimodal End-to-End Autonomous Driving, arXiv preprint arXiv:1906.03199 (2019).*

[17] *Zhang Y, O. Morel, M. Blanchon, R. Seulin, M. Rastgoo, D. Sidibé, Exploration of Deep Learning-based Multimodal Fusion for Semantic Road Scene Segmentation, in: VISIGRAPP 2019 - Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, volume 5, 2019, pp. 336–343. 1115 doi:10.5220/0007360403360343.*