# Multi-View Capsule Network

Jianwei Liu , Xi-hao Ding, Runkun Lu, Yuan-feng Lian, Dian-zhong Wang, Xionglin Luo
Contact Author: (Jian-wei Liu 2236677012@qq.com)

Multi-view learning attempts to generate a model with a better performance by exploiting information among multi-view data. Most existing approaches only focus on either consistency or complementarity principle, and learn representations (or features) of the multi-view data. In this paper, to utilize both complementarity and consistency simultaneously, and explore the potential of deep learning in multi-view learning, we propose a novel supervised multi-view learning algorithm, called multi-view capsule network (MVCapsNet), which extracts a feature matrix of all views by a group of encoders, and obtains a classification matrix fusing common and special information of multiple views. Extensive experiments conducted on eight real-world datasets have demonstrated the effectiveness of our proposed method, and show its superiority over several state-of-the-art baseline methods.

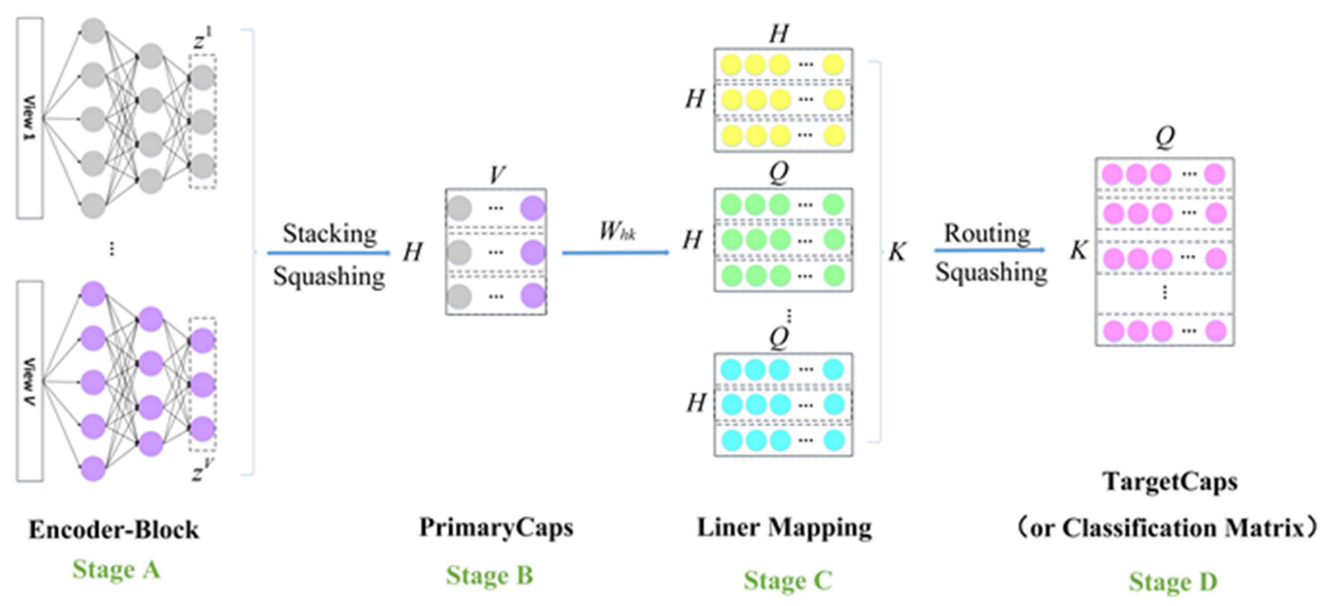


**Fig. 1.** We illustrate a V-views problem in this figure. First, Encoder-Block parallel connects the output of each encoder in Encoder-Block, obtain a feature matrix with the shape of . Then PrimaryCaps outputs a squashed feature matrix, which will be further linear mapped, routed, and squashed. Finally, TargetCaps outputs a classification matrix.

**Table 1.** Accuracy of different methods

| Method | ACC(%) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Leaves | Reuters | YaleFace | BBC | Cornell | Texas | Washington | Wisconsin |
| GNMF | 95.0±0 | 40.8±1.2 | 50.0±2.5 | 38.0±1.5 | 41.0±1.8 | 57.9±1.8 | 69.6±2.2 | 52.8±1.4 |
| MultiNMF | 95.0±0 | 52.7±0.2 | 64.2±4.2 | 73.1±0.2 | 49.7±7.7 | 68.7±3.4 | 59.3±2.6 | 50.3±3.5 |
| MVCC | 100±0 | 54.4±1.9 | 33.3±6.9 | 95.8±2.6 | 60.8±5.0 | 64.7±5.5 | 62.8±3.8 | 64.3±2.7 |
| DICS | 97.9±2.5 | 70.3±4.0 | 89.1±3.2 | 90.2±2.4 | 72.8±6.1 | 81.6±4.0 | 77.4±6.0 | 85.1±4.5 |
| Max-Pooling | 100±0 | 90.0±4.6 | 71.2±3.4 | 80.5±7.6 | 71.3±8.7 | 74.7±5.2 | 67.5±8.2 | 86.2±7.7 |
| Mean-Pooling | 100±0 | 90.6±5.3 | 71.2±4.3 | 83.3±6.7 | 70.9±5.5 | 76.6±4.0 | 70.0±4.9 | 84.8±5.2 |
| Weighted Sum | 100±0 | 92.9±4.8 | 72.8±4.7 | 87.2±4.5 | 72.5±13 | 76.3±4.9 | 66.9±7.8 | 86.3±4.3 |
| **MVCapsNet** | **100±0** | **76.8±3.2** | **95.8±2.3** | **96.2±1.9** | **76.3±3.2** | **84.0±5.9** | **91.9±7.4** | **87.4±3.8** |

To our best knowledge, this work represents the first attempt to extend capsule net to multi-view learning scenarios. As a result, our contributions we summarize are shown as follows:

(1) different from concatenation, MVCapsNet vertically stack feature vectors extracted by encoders to product a feature matrix, which capture the non-linear relationship and real underlying properties in multi-view dataset;

(2) MVCapsNet is a supervised multi-view deep learning algorithm utilizing both consistency and complementarity of multiple views by capsule layers and dynamic routing mechanism, where multiple views' encoders consider the consistency, and the linear mapping of routing process considers the complementarity;

(3) the mini-batch method can be used to train the model parameters of MVCapsNet, thus unlike traditional multi-view learning methods that based on matrix factorization, MVCapsNet can be applied to large scale data sets;

(4) MVCapsNet avoids the slowly training process because that the number of views is always limited in a small range, and comparing with traditional Capsule Net, the number of parameters in routing process decreased from 106 to 104.

(5) we also build other baselines deep networks to further analyze MVCapsNet's performance, which explore complementary by mean-pooling, max-pooling and weighted summation. Experimental results show that MVCapsNet outperforms the baselines and factorization base methods, finally achieves better accuracy on all datasets, particularly, comparing with the second best algorithm, the accuracy of our approach is increased by 14.5% on Washington dataset.