시퀀스 생성을 통한 제로샷 행동 인식 해결방법

이제욱, 박성호, 변혜란
연세대학교

(hooraid, qkrtjdgh18, hrbyun)@yonsei.ac.kr

Multiple Features Generative Approach to Zero-Shot Learning

Lee Jewook, Park Sungho, Byun Hyeran
Yonsei Univ.

(hooraid, qkrtjdgh18, hrbyun)@yonsei.ac.kr

요 약

Generating an averaged feature vectors on the unseen video is intuitive way for zero-shot learning, but it crushes temporal organization in a video. To address this problem, we propose the new paradigm, which generate not a averaged feature, but features for each video. Our networks is composed of three modules, attribute encoder, generator and two discriminator.

Ⅰ. 서 론

Recently, Zero-shot Learning has been a popular topic in computer science due to it is a more practical scenario. In detail, the Zero-shot Learning scenario is to classify unseen classes which are not included in the training dataset. In zero-shot image classification, feature generating for training unseen classes is considered a powerful but simple approach.

Similarly, in zero-shot action recognition, existing methods[1,2,3] synthesize a single feature even though the video has temporal information. So we propose a new novel approach to generate multiple features for a video. Our network is composed of three modules, attribute encoder, generator, and two discriminators.

Ⅱ. 본론

For generating multiple features from a single latent vector and attribute, we first design an attribute encoder. By using Recurrent Neural Networks, the encoder can increase the number of features. Also, For ensuring diversity of latent space, we utilize the re-parametrization trick with KL divergence. It helps to augment condition and latent by sampling from the gaussian distribution.

Our network is based on Generative Adversarial Networks(GAN). The proposed network has a generator and discriminator. It is enough for existing methods because they generate a single feature from a condition. However, we aim to generate multiple features. Therefore, the network should consider the nature of temporal flow among generated features. For achieving the goal, we propose a new discriminator to scrutinize whole features.

We evaluate our proposed networks by the GZSL metric[5], which is the harmonic mean of the accuracy of seen actions and unseen actions. We set a baseline as f-CLSWGAN[6] with adaption from image classification to action recognition. This represents the model which generates an averaged feature. Though baseline records 31.4±3.2, our full model, which contains the attribute encoder and two discriminators, achieves the superior performance by 35.1±3.2.

Ⅲ. 결론

Through several experiments, we verified effectiveness of a sequence generative approach. Also, we conclude that sequence generating with enough fidelity could outperform than averaged feature generating.

ACKNOWLEDGMENT

"This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2021-2016-0-00464) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation)"

참 고 문 헌

[1] Christoph H. Lampert, Hannes Nickisch, and Stefan Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 951- 958, 2009.

[2] Zeynep Akata, Scott E. Reed, Daniel Walter, Honglak Lee, and Bernt Schiele. Evaluation of output embeddings for fine-grained image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2927- 2936, 2015.

[3] Long Chen, Hanwang Zhang, Jun Xiao, Wei Liu, and Shih-Fu Chang. Zero-shot visual recognition using semantics-preserving adversarial embedding networks. In The IEEE Conference on Computer Vision and Pattern Recognition(CVPR), June 2018.

[4] Yashas Annadani and Soma Biswas. Preserving semantic relations for zero-shot learning. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June2018.

[5] Y. Xian, B. Schiele, Z. Akata, Zero-shot learning - the good, the bad and the ugly, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3077- 30

[6] Y. Xian, T. Lorenz, B. Schiele, Z. Akata, Feature generating networks for zero-shot learning, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018