FVQA: Fact-Based Visual Question Answering [1]

1. They contributed to the development of answering more complex questions visually
2. Concluded that fact based support answers are critical for image querying
3. Their vision learning comes from mapping of questions to Knowledge Bases (which hold facts based support answers)
4. Visual Question Answer (VQA) is limited, when asked question outside its scope.
1. Knowledge Bases have a structure of a triple, 2 arguments (objects) and its relation.
2. That from an RNN approach it is impossible to tell if correct answers are based the image information or just by a particular training set or pattern.
3. From image set, extracted and identified objects, scenes and actions used to make questions
4. The KB query are visual concepts and also relationships, which give possible supporting facts
1. Using GPS for the analyst of image can increase speed of extraction of objects in images.
2. The KBs are stored as a graph which can be passed to GPU for quicker queries.
3. With the system has to use both image information and FVQA and decide which it needs to use to answer the question, CPU is used heavily on the computation comparisons.

Deep Bimodal Regression of Apparent Personality Traits from Short Video Sequences [2]

1. The Deep bimodal regression was to evaluate “Big Five” model which consist of openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism.
2. The evaluation of the personalities was split into audio and visual deep learning.
3. Convolution Neural Networks are good for deep learning on image content and context.
4. CNN was used for the deep learning of the audio modal.
1. For the visual learning, the image set extracted was at a rate of 6fps (approx. 100 frames)
2. Descriptor Aggregation Network (DAN) is a modified CNN which adds an averaging and max pooling layer.
3. Due to DANs not being fully connected layers, it has improved performance over traditional CNN.
4. Comparable models, ResNet and VGG could not focus attention on a number of the human beings from the sample images.
1. GPUs were used in the training for audio and visual training
2. GPUs can help the accelerate the extracting which can increase the no. of images sampled
3. With GPU being used there will be less time spent training and thus reducing energy consumed.

[1] Peng Wang , Qi Wu , Chunhua Shen , Anthony Dick, and Anton van den Hengel. “FVQA: Fact-Based Visual Question Answering” in IEEE Transactions on Pattern Analysis and Machine Intelligences, VOL. 40, NO. 10, October 2018
[2] Xiu-Shen Wei , Chen-Lin Zhang, Hao Zhang , and Jianxin Wu , Member, IEEE. “Deep Bimodal Regression of Apparent Personality Traits from Short Video Sequences” in IEEE Transactions on Affective Computing, VOL. 9, NO. 3, July-September 2018