

Papers to read:

- Convolutional neural network architecture for geometric matching:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Rocco_Convolutional_Neural_Network_CVPR_2017_paper.pdf
- Dynamic Edge-Conditioned Filters in Convolutional Neural Networks on Graphs:
http://openaccess.thecvf.com/content_cvpr_2017/html/Simonovsky_Dynamic_Edge-Conditioned_Filters_CVPR_2017_paper.html
- YOLO-9000: <https://arxiv.org/pdf/1612.08242.pdf>
- Deep Reinforcement Learning-based Image Captioning with Embedding Reward:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Ren_Deep_Reinforcement_Learning-Based_CVPR_2017_paper.pdf
- Interpretable Structure-Evolving LSTM:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Liang_Interpretable_Structure-Evolving_LSTM_CVPR_2017_paper.pdf
- Inverse Compositional Spatial Transformer Networks:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Lin_Inverse_Compositional_Spatial_CVPR_2017_paper.pdf
- Learning from Simulated and Unsupervised Images through Adversarial Training:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Shrivastava_Learning_From_Simulated_CVPR_2017_paper.pdf
- Densely Connected Convolutional Networks:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.pdf
- Video Propagation Networks:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Jampani_Video_Propagation_Networks_CVPR_2017_paper.pdf
- Scene Graph Generation by Iterative Message Passing:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Xu_Scene_Graph_Generation_CVPR_2017_paper.pdf
- DSAC - Differentiable RANSAC for Camera Localization:
<https://arxiv.org/pdf/1611.05705.pdf>
- From Red Wine to Red Tomato: Composition with Context:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Misra_From_Red_Wine_CVPR_2017_paper.pdf
- Hierarchical Boundary-Aware Neural Encoder for Video Captioning:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Baraldi_Hierarchical_Boundary-Aware_Neural_CVPR_2017_paper.pdf

Other interesting papers:

- Synthesizing Dynamic Patterns by Spatial-Temporal Generative ConvNet:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Xie_Synthesizing_Dynamic_Patterns_CVPR_2017_paper.pdf

- CDC: Convolutional-De-Convolutional Networks for Precise Temporal Action Localization in Untrimmed Videos:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Shou_CDC_Convolutional-De-Convolutional_Networks_CVPR_2017_paper.pdf
- Predictive-Corrective Networks for Action Detection:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Dave_Predictive-Corrective_Networks_for_CVPR_2017_paper.pdf
- Temporal Convolutional Networks for Action Segmentation and Detection:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Lea_Temporal_Convolutional_Networks_CVPR_2017_paper.pdf
- Temporal Attention-Gated Model for Robust Sequence Classification:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Pei_Temporal_Attention-Gated_Model_CVPR_2017_paper.pdf
- Feature Pyramid Networks for Object Detection:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Lin_Feature_Pyramid_Networks_CVPR_2017_paper.pdf
- Borrowing Treasures from the Wealthy: Deep Transfer Learning through Selective Joint Fine-Tuning:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Ge_Borrowing_Treasures_From_CVPR_2017_paper.pdf
- Deep Image Matting:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Xu_Deep_Image_Matting_CVPR_2017_paper.pdf
- TGIF-QA: Toward Spatio-Temporal Reasoning in Visual Question Answering:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Jang_TGIF-QA_Toward_Spatio-Temporal_CVPR_2017_paper.pdf
- Enhancing Video Summarization via Vision-Language Embedding:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Plummer_Enhancing_Video_Summarization_CVPR_2017_paper.pdf
- Semantic Compositional Networks for Visual Captioning:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Gan_Semantic_Compositional_Networks_CVPR_2017_paper.pdf
- Query-Focused Video Summarization: Dataset, Evaluation, and A Memory Network Based Approach:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Sharghi_Query-Focused_Video_Summarization_CVPR_2017_paper.pdf
- Dense Captioning with Joint Inference and Visual Context:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Yang_Dense_Captioning_With_CVPR_2017_paper.pdf
- Deep Feature Flow for Video Recognition:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Zhu_Deep_Feature_Flow_CVPR_2017_paper.pdf

- Real-Time Video Super-Resolution with Spatio-Temporal Networks and Motion Compensation:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Caballero_Real-Time_Video_Super-Resolution_CVPR_2017_paper.pdf
- AnchorNet: A Weakly Supervised Network to Learn Geometry-sensitive Features For Semantic Matching:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Novotny_AnchorNet_A_Weakly_CVPR_2017_paper.pdf
- Semi-Supervised Deep Learning for Monocular Depth Map Prediction:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Kuznietsov_Semi-Supervised_Deep_Learning_CVPR_2017_paper.pdf
- Learning Cross-modal Embeddings
 - for Cooking Recipes and Food Images:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Salvador_Learning_Cross-Modal_Embeddings_CVPR_2017_paper.pdf
 - for Robust Pedestrian Detection:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Xu_Learning_Cross-Modal_Deep_CVPR_2017_paper.pdf
- High-Resolution Image Inpainting using Multi-Scale Neural Patch Synthesis:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Yang_High-Resolution_Image_Inpainting_CVPR_2017_paper.pdf
- HPatches: A benchmark and evaluation of handcrafted and learned local descriptors:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Balntas_HPatches_A_Benchmark_CVPR_2017_paper.pdf
- Not All Pixels Are Equal: Difficulty-Aware Semantic Segmentation via Deep Layer Cascade:
http://openaccess.thecvf.com/content_cvpr_2017/papers/Li_Not_All_Pixels_CVPR_2017_paper.pdf