TITLE: It's Not About the Journey; It's About the Destination: Following Soft Paths Under Question-Guidance for Visual Reasoning
http://openaccess.thecvf.com/content_CVPR_2019/html/Haurilet_Its_Not_About_the_Journey_Its_About_the_Destination_Following_CVPR_2019_paper.html
AUTHORS: Monica Haurilet, Alina Roitberg, Rainer Stiefelhagen
HIGHLIGHT: We present a new model for Visual Reasoning, aimed at capturing the interplay among individual objects in the image represented as a scene graph.

TITLE: Actively Seeking and Learning From Live Data
http://openaccess.thecvf.com/content_CVPR_2019/html/Teney_Actively_Seeking_and_Learning_From_Live_Data_CVPR_2019_paper.html
AUTHORS: Damien Teney, Anton van den Hengel
HIGHLIGHT: The approach we propose is a step toward overcoming this limitation by searching for the information required at test time.

TITLE: Improving Referring Expression Grounding With Cross-Modal Attention-Guided Erasing
http://openaccess.thecvf.com/content_CVPR_2019/html/Liu_Improving_Referring_Expression_Grounding_With_Cross-Modal_Attention-Guided_Erasing_CVPR_2019_paper.html
AUTHORS: Xihui Liu, Zihao Wang, Jing Shao, Xiaogang Wang, Hongsheng Li
HIGHLIGHT: To tackle this issue, we design a novel cross-modal attention-guided erasing approach, where we discard the most dominant information from either textual or visual domains to generate difficult training samples online, and to drive the model to discover complementary textual-visual correspondences.

TITLE: Neighbourhood Watch: Referring Expression Comprehension via Language-Guided Graph Attention Networks
http://openaccess.thecvf.com/content_CVPR_2019/html/Wang_Neighbourhood_Watch_Referring_Expression_Comprehension_via_Language-Guided_Graph_Attention_Networks_CVPR_2019_paper.html
AUTHORS: Peng Wang, Qi Wu, Jiewei Cao, Chunhua Shen, Lianli Gao, Anton van den Hengel
HIGHLIGHT: To capture and exploit this important information we propose a graph-based, language-guided attention mechanism.

TITLE: Scene Graph Generation With External Knowledge and Image Reconstruction
http://openaccess.thecvf.com/content_CVPR_2019/html/Gu_Scene_Graph_Generation_With_External_Knowledge_and_Image_Reconstruction_CVPR_2019_paper.html
AUTHORS: Jiuxiang Gu, Handong Zhao, Zhe Lin, Sheng Li, Jianfei Cai, Mingyang Ling
HIGHLIGHT: In this paper, we propose a novel scene graph generation algorithm with external knowledge and image reconstruction loss to overcome these dataset issues.

TITLE: Polysemous Visual-Semantic Embedding for Cross-Modal Retrieval
http://openaccess.thecvf.com/content_CVPR_2019/html/Song_Polysemous_Visual-Semantic_Embedding_for_Cross-Modal_Retrieval_CVPR_2019_paper.html
AUTHORS: Yale Song, Mohammad Soleymani
HIGHLIGHT: In this work, we introduce Polysemous Instance Embedding Networks (PIE-Nets) that compute multiple and diverse representations of an instance by combining global context with locally-guided features via multi-head self-attention and residual learning.

TITLE: MUREL: Multimodal Relational Reasoning for Visual Question Answering
http://openaccess.thecvf.com/content_CVPR_2019/html/Cadene_MUREL_Multimodal_Relational_Reasoning_for_Visual_Question_Answering_CVPR_2019_paper.html
AUTHORS: Remi Cadene, Hedi Ben-younes, Matthieu Cord, Nicolas Thome
HIGHLIGHT: In this paper, we propose MuRel, a multimodal relational network which is learned end-to-end to reason over real images.

TITLE: Heterogeneous Memory Enhanced Multimodal Attention Model for Video Question Answering
http://openaccess.thecvf.com/content_CVPR_2019/html/Fan_Heterogeneous_Memory_Enhanced_Multimodal_Attention_Model_for_Video_Question_Answering_CVPR_2019_paper.html
AUTHORS: Chenyou Fan, Xiaofan Zhang, Shu Zhang, Wensheng Wang, Chi Zhang, Heng Huang
HIGHLIGHT: In this paper, we propose a novel end-to-end trainable Video Question Answering (VideoQA) framework with three major components: 1) a new heterogeneous memory which can effectively learn global context information from appearance and motion features; 2) a redesigned question memory which helps understand the complex semantics of question and highlights queried subjects; and 3) a new multimodal fusion layer which performs multi-step reasoning by attending to relevant visual and textual hints with self-updated attention.

TITLE: Information Maximizing Visual Question Generation
http://openaccess.thecvf.com/content_CVPR_2019/html/Krishna_Information_Maximizing_Visual_Question_Generation_CVPR_2019_paper.html
AUTHORS: Ranjay Krishna, Michael Bernstein, Li Fei-Fei
HIGHLIGHT: To overcome the non-differentiability of discrete natural language tokens, we introduce a variational continuous latent space onto which the expected answers project.

TITLE: Learning to Detect Human-Object Interactions With Knowledge
http://openaccess.thecvf.com/content_CVPR_2019/html/Xu_Learning_to_Detect_Human-Object_Interactions_With_Knowledge_CVPR_2019_paper.html
AUTHORS: Bingjie Xu, Yongkang Wong, Junnan Li, Qi Zhao, Mohan S. Kankanhalli
HIGHLIGHT: In this work, we focus on detecting human-object interactions (HOIs) in images, an essential step towards deeper scene understanding.

TITLE: Learning Words by Drawing Images
http://openaccess.thecvf.com/content_CVPR_2019/html/Suris_Learning_Words_by_Drawing_Images_CVPR_2019_paper.html
AUTHORS: Didac Suris, Adria Recasens, David Bau, David Harwath, James Glass, Antonio Torralba
HIGHLIGHT: We propose a framework for learning through drawing.

TITLE: Factor Graph Attention
http://openaccess.thecvf.com/content_CVPR_2019/html/Schwartz_Factor_Graph_Attention_CVPR_2019_paper.html
AUTHORS: Idan Schwartz, Seunghak Yu, Tamir Hazan, Alexander G. Schwing
HIGHLIGHT: We address this issue and develop a general attention mechanism for visual dialog which operates on any number of data utilities.

TITLE: Unsupervised Image Captioning
http://openaccess.thecvf.com/content_CVPR_2019/html/Feng_Unsupervised_Image_Captioning_CVPR_2019_paper.html
AUTHORS: Yang Feng, Lin Ma, Wei Liu, Jiebo Luo
HIGHLIGHT: In this paper, we make the first attempt to train an image captioning model in an unsupervised manner.

TITLE: Exact Adversarial Attack to Image Captioning via Structured Output Learning With Latent Variables
http://openaccess.thecvf.com/content_CVPR_2019/html/Xu_Exact_Adversarial_Attack_to_Image_Captioning_via_Structured_Output_Learning_CVPR_2019_paper.html
AUTHORS: Yan Xu, Baoyuan Wu, Fumin Shen, Yanbo Fan, Yong Zhang, Heng Tao Shen, Wei Liu
HIGHLIGHT: In this work, we study the robustness of a CNN+RNN based image captioning system being subjected to adversarial noises.

TITLE: Cross-Modal Relationship Inference for Grounding Referring Expressions
http://openaccess.thecvf.com/content_CVPR_2019/html/Yang_Cross-Modal_Relationship_Inference_for_Grounding_Referring_Expressions_CVPR_2019_paper.html
AUTHORS: Sibei Yang, Guanbin Li, Yizhou Yu
HIGHLIGHT: In this paper, we propose a Cross-Modal Relationship Extractor (CMRE) to adaptively highlight objects and relationships, that have connections with a given expression, with a cross-modal attention mechanism, and represent the extracted information as a language-guided visual relation graph.

TITLE: What's to Know? Uncertainty as a Guide to Asking Goal-Oriented Questions
http://openaccess.thecvf.com/content_CVPR_2019/html/Abbasnejad_Whats_to_Know_Uncertainty_as_a_Guide_to_Asking_Goal-Oriented_CVPR_2019_paper.html
AUTHORS: Ehsan Abbasnejad, Qi Wu, Qinfeng Shi, Anton van den Hengel
HIGHLIGHT: We propose a solution to this problem based on a Bayesian model of the uncertainty in the implicit model maintained by the visual dialogue agent, and in the function used to select an appropriate output.

TITLE: Iterative Alignment Network for Continuous Sign Language Recognition
http://openaccess.thecvf.com/content_CVPR_2019/html/Pu_Iterative_Alignment_Network_for_Continuous_Sign_Language_Recognition_CVPR_2019_paper.html
AUTHORS: Junfu Pu, Wengang Zhou, Houqiang Li
HIGHLIGHT: In this paper, we propose an alignment network with iterative optimization for weakly supervised continuous sign language recognition.

TITLE: Neural Sequential Phrase Grounding (SeqGROUND)

http://openaccess.thecvf.com/content_CVPR_2019/html/Dogan_Neural_Sequential_Phrase_Grounding_SeqGROUND_CVPR_2019_paper.html
AUTHORS: Pelin Dogan, Leonid Sigal, Markus Gross
HIGHLIGHT: We propose an end-to-end approach for phrase grounding in images.

TITLE: CLEVR-Ref+: Diagnosing Visual Reasoning With Referring Expressions
http://openaccess.thecvf.com/content_CVPR_2019/html/Liu_CLEVR-Ref_Diagnosing_Visual_Reasoning_With_Referring_Expressions_CVPR_2019_paper.html
AUTHORS: Runtao Liu, Chenxi Liu, Yutong Bai, Alan L. Yuille
HIGHLIGHT: In particular, we present two interesting and important findings using IEP-Ref: (1) the module trained to transform feature maps into segmentation masks can be attached to any intermediate module to reveal the entire reasoning process step-by-step; (2) even if all training data has at least one object referred, IEP-Ref can correctly predict no-foreground when presented with false-premise referring expressions.

TITLE: Describing Like Humans: On Diversity in Image Captioning
http://openaccess.thecvf.com/content_CVPR_2019/html/Wang_Describing_Like_Humans_On_Diversity_in_Image_Captioning_CVPR_2019_paper.html
AUTHORS: Qingzhong Wang, Antoni B. Chan
HIGHLIGHT: In this paper, we proposed a new metric for measuring the diversity of image captions, which is derived from latent semantic analysis and kernelized to use CIDEr similarity.

TITLE: MSCap: Multi-Style Image Captioning With Unpaired Stylized Text
http://openaccess.thecvf.com/content_CVPR_2019/html/Guo_MSCap_Multi-Style_Image_Captioning_With_Unpaired_Stylized_Text_CVPR_2019_paper.html
AUTHORS: Longteng Guo, Jing Liu, Peng Yao, Jiangwei Li, Hanqing Lu
HIGHLIGHT: In this paper, we propose an adversarial learning network for the task of multi-style image captioning (MSCap) with a standard factual image caption dataset and a multi-stylized language corpus without paired images.

TITLE: Context and Attribute Grounded Dense Captioning
http://openaccess.thecvf.com/content_CVPR_2019/html/Yin_Context_and_Attribute_Grounded_Dense_Captioning_CVPR_2019_paper.html
AUTHORS: Guojun Yin, Lu Sheng, Bin Liu, Nenghai Yu, Xiaogang Wang, Jing Shao
HIGHLIGHT: In this work, we investigate contextual reasoning based on multi-scale message propagations from the neighboring contents to the target ROIs.

TITLE: Spot and Learn: A Maximum-Entropy Patch Sampler for Few-Shot Image Classification
http://openaccess.thecvf.com/content_CVPR_2019/html/Chu_Spot_and_Learn_A_Maximum-Entropy_Patch_Sampler_for_Few-Shot_Image_CVPR_2019_paper.html
AUTHORS: Wen-Hsuan Chu, Yu-Jhe Li, Jing-Cheng Chang, Yu-Chiang Frank Wang
HIGHLIGHT: In this work, we propose a sampling method that de-correlates an image based on maximum entropy reinforcement learning, and extracts varying sequences of patches on every forward-pass with discriminative information observed.

TITLE: Interpreting CNNs via Decision Trees
http://openaccess.thecvf.com/content_CVPR_2019/html/Zhang_Interpreting_CNNs_via_Decision_Trees_CVPR_2019_paper.html
AUTHORS: Quanshi Zhang, Yu Yang, Haotian Ma, Ying Nian Wu
HIGHLIGHT: This paper aims to quantitatively explain the rationales of each prediction that is made by a pre-trained convolutional neural network (CNN).

TITLE: Dense Relational Captioning: Triple-Stream Networks for Relationship-Based Captioning
http://openaccess.thecvf.com/content_CVPR_2019/html/Kim_Dense_Relational_Captioning_Triple-Stream_Networks_for_Relationship-Based_Captioning_CVPR_2019_paper.html
AUTHORS: Dong-Jin Kim, Jinsoo Choi, Tae-Hyun Oh, In So Kweon
HIGHLIGHT: Our goal in this work is to train an image captioning model that generates more dense and informative captions.

TITLE: Deep Modular Co-Attention Networks for Visual Question Answering
http://openaccess.thecvf.com/content_CVPR_2019/html/Yu_Deep_Modular_Co-Attention_Networks_for_Visual_Question_Answering_CVPR_2019_paper.html
AUTHORS: Zhou Yu, Jun Yu, Yuhao Cui, Dacheng Tao, Qi Tian
HIGHLIGHT: In this paper, we propose a deep Modular Co-Attention Network (MCAN) that consists of Modular Co-Attention (MCA) layers cascaded in depth.

TITLE: Synthesizing Environment-Aware Activities via Activity Sketches
http://openaccess.thecvf.com/content_CVPR_2019/html/Liao_Synthesizing_Environment-Aware_Activities_via_Activity_Sketches_CVPR_2019_paper.html
AUTHORS: Yuan-Hong Liao, Xavier Puig, Marko Boben, Antonio Torralba, Sanja Fidler
HIGHLIGHT: In this work, we address the problem: environment-aware program generation.

TITLE: Self-Critical N-Step Training for Image Captioning
http://openaccess.thecvf.com/content_CVPR_2019/html/Gao_Self-Critical_N-Step_Training_for_Image_Captioning_CVPR_2019_paper.html
AUTHORS: Junlong Gao, Shiqi Wang, Shanshe Wang, Siwei Ma, Wen Gao
HIGHLIGHT: In this paper, we estimate state value without using a parametrized value estimator.

TITLE: Multi-Target Embodied Question Answering
http://openaccess.thecvf.com/content_CVPR_2019/html/Yu_Multi-Target_Embodied_Question_Answering_CVPR_2019_paper.html
AUTHORS: Licheng Yu, Xinlei Chen, Georgia Gkioxari, Mohit Bansal, Tamara L. Berg, Dhruv Batra
HIGHLIGHT: We present a generalization of EQA -- Multi-Target EQA (MT-EQA).

TITLE: Visual Question Answering as Reading Comprehension
http://openaccess.thecvf.com/content_CVPR_2019/html/Li_Visual_Question_Answering_as_Reading_Comprehension_CVPR_2019_paper.html
AUTHORS: Hui Li, Peng Wang, Chunhua Shen, Anton van den Hengel
HIGHLIGHT: In contrast to struggling on multimodal feature fusion, in this paper, we propose to unify all the input information by natural language so as to convert VQA into a machine reading comprehension problem.

TITLE: StoryGAN: A Sequential Conditional GAN for Story Visualization
http://openaccess.thecvf.com/content_CVPR_2019/html/Li_StoryGAN_A_Sequential_Conditional_GAN_for_Story_Visualization_CVPR_2019_paper.html
AUTHORS: Yitong Li, Zhe Gan, Yelong Shen, Jingjing Liu, Yu Cheng, Yuexin Wu, Lawrence Carin, David Carlson, Jianfeng Gao
HIGHLIGHT: In this work, we propose a new task called Story Visualization.

TITLE: Grounded Video Description
http://openaccess.thecvf.com/content_CVPR_2019/html/Zhou_Grounded_Video_Description_CVPR_2019_paper.html
AUTHORS: Luowei Zhou, Yannis Kalantidis, Xinlei Chen, Jason J. Corso, Marcus Rohrbach
HIGHLIGHT: In this work, we explicitly link the sentence to the evidence in the video by annotating each noun phrase in a sentence with the corresponding bounding box in one of the frames of a video.

TITLE: Streamlined Dense Video Captioning
http://openaccess.thecvf.com/content_CVPR_2019/html/Mun_Streamlined_Dense_Video_Captioning_CVPR_2019_paper.html
AUTHORS: Jonghwan Mun, Linjie Yang, Zhou Ren, Ning Xu, Bohyung Han
HIGHLIGHT: To tackle this challenge, we propose a novel dense video captioning framework, which models temporal dependency across events in a video explicitly and leverages visual and linguistic context from prior events for coherent storytelling.

TITLE: Adversarial Inference for Multi-Sentence Video Description
http://openaccess.thecvf.com/content_CVPR_2019/html/Park_Adversarial_Inference_for_Multi-Sentence_Video_Description_CVPR_2019_paper.html
AUTHORS: Jae Sung Park, Marcus Rohrbach, Trevor Darrell, Anna Rohrbach
HIGHLIGHT: In this work, we instead propose to apply adversarial techniques during inference, designing a discriminator which encourages better multi-sentence video description.

TITLE: Unified Visual-Semantic Embeddings: Bridging Vision and Language With Structured Meaning Representations
http://openaccess.thecvf.com/content_CVPR_2019/html/Wu_Unified_Visual-Semantic_Embeddings_Bridging_Vision_and_Language_With_Structured_Meaning_CVPR_2019_paper.html
AUTHORS: Hao Wu, Jiayuan Mao, Yufeng Zhang, Yuning Jiang, Lei Li, Weiwei Sun, Wei-Ying Ma
HIGHLIGHT: We propose the Unified Visual-Semantic Embeddings (Unified VSE) for learning a joint space of visual representation and textual semantics.

TITLE: Learning to Compose Dynamic Tree Structures for Visual Contexts
http://openaccess.thecvf.com/content_CVPR_2019/html/Tang_Learning_to_Compose_Dynamic_Tree_Structures_for_Visual_Contexts_CVPR_2019_paper.html
AUTHORS: Kaihua Tang, Hanwang Zhang, Baoyuan Wu, Wenhan Luo, Wei Liu

HIGHLIGHT: We propose to compose dynamic tree structures that place the objects in an image into a visual context, helping visual reasoning tasks such as scene graph generation and visual Q&A.

TITLE: Reinforced Cross-Modal Matching and Self-Supervised Imitation Learning for Vision-Language Navigation
http://openaccess.thecvf.com/content_CVPR_2019/html/Wang_Reinforced_Cross-Modal_Matching_and_Self-Supervised_Imitation_Learning_for_Vision-Language_Navigation_CVPR_2019_paper.html
AUTHORS: Xin Wang, Qiuyuan Huang, Asli Celikyilmaz, Jianfeng Gao, Dinghan Shen, Yuan-Fang Wang, William Yang Wang, Lei Zhang
HIGHLIGHT: In this paper, we study how to address three critical challenges for this task: the cross-modal grounding, the ill-posed feedback, and the generalization problems.

TITLE: Dynamic Fusion With Intra- and Inter-Modality Attention Flow for Visual Question Answering
http://openaccess.thecvf.com/content_CVPR_2019/html/Gao_Dynamic_Fusion_With_Intra-_and_Inter-Modality_Attention_Flow_for_Visual_CVPR_2019_paper.html
AUTHORS: Peng Gao, Zhengkai Jiang, Haoxuan You, Pan Lu, Steven C. H. Hoi, Xiaogang Wang, Hongsheng Li
HIGHLIGHT: We propose a novel method of dynamically fuse multi-modal features with intra- and inter-modality information flow, which alternatively pass dynamic information between and across the visual and language modalities.

TITLE: Cycle-Consistency for Robust Visual Question Answering
http://openaccess.thecvf.com/content_CVPR_2019/html/Shah_Cycle-Consistency_for_Robust_Visual_Question_Answering_CVPR_2019_paper.html
AUTHORS: Meet Shah, Xinlei Chen, Marcus Rohrbach, Devi Parikh
HIGHLIGHT: As a step towards improving robustness of VQA models, we propose a model-agnostic framework that exploits cycle consistency.

TITLE: Embodied Question Answering in Photorealistic Environments With Point Cloud Perception
http://openaccess.thecvf.com/content_CVPR_2019/html/Wijmans_Embodied_Question_Answering_in_Photorealistic_Environments_With_Point_Cloud_Perception_CVPR_2019_paper.html
AUTHORS: Erik Wijmans, Samyak Datta, Oleksandr Maksymets, Abhishek Das, Georgia Gkioxari, Stefan Lee, Irfan Essa, Devi Parikh, Dhruv Batra
HIGHLIGHT: We find that two seemingly naive navigation baselines, forward-only and random, are strong navigators and challenging to outperform, due to the specific choice of the evaluation setting presented by [1].

TITLE: Reasoning Visual Dialogs With Structural and Partial Observations
http://openaccess.thecvf.com/content_CVPR_2019/html/Zheng_Reasoning_Visual_Dialogs_With_Structural_and_Partial_Observations_CVPR_2019_paper.html
AUTHORS: Zilong Zheng, Wenguan Wang, Siyuan Qi, Song-Chun Zhu
HIGHLIGHT: We propose a novel model to address the task of Visual Dialog which exhibits complex dialog structures.

TITLE: Recursive Visual Attention in Visual Dialog
http://openaccess.thecvf.com/content_CVPR_2019/html/Niu_Recursive_Visual_Attention_in_Visual_Dialog_CVPR_2019_paper.html
AUTHORS: Yulei Niu, Hanwang Zhang, Manli Zhang, Jianhong Zhang, Zhiwu Lu, Ji-Rong Wen
HIGHLIGHT: In this work, to resolve the visual co-reference for visual dialog, we propose a novel attention mechanism called Recursive Visual Attention (RvA).

TITLE: Two Body Problem: Collaborative Visual Task Completion
http://openaccess.thecvf.com/content_CVPR_2019/html/Jain_Two_Body_Problem_Collaborative_Visual_Task_Completion_CVPR_2019_paper.html
AUTHORS: Unnat Jain, Luca Weihs, Eric Kolve, Mohammad Rastegari, Svetlana Lazebnik, Ali Farhadi, Alexander G. Schwing, Aniruddha Kembhavi
HIGHLIGHT: In this paper we study the problem of learning to collaborate directly from pixels in AI2-THOR and demonstrate the benefits of explicit and implicit modes of communication to perform visual tasks.

TITLE: GQA: A New Dataset for Real-World Visual Reasoning and Compositional Question Answering
http://openaccess.thecvf.com/content_CVPR_2019/html/Hudson_GQA_A_New_Dataset_for_Real-World_Visual_Reasoning_and_Compositional_CVPR_2019_paper.html
AUTHORS: Drew A. Hudson, Christopher D. Manning
HIGHLIGHT: We introduce GQA, a new dataset for real-world visual reasoning and compositional question answering, seeking to address key shortcomings of previous VQA datasets.

TITLE: Text2Scene: Generating Compositional Scenes From Textual Descriptions
http://openaccess.thecvf.com/content_CVPR_2019/html/Tan_Text2Scene_Generating_Compositional_Scenes_From_Textual_Descriptions_CVPR_2019_paper.html
AUTHORS: Fuwen Tan, Song Feng, Vicente Ordonez
HIGHLIGHT: In this paper, we propose Text2Scene, a model that generates various forms of compositional scene representations from natural language descriptions.

TITLE: From Recognition to Cognition: Visual Commonsense Reasoning
http://openaccess.thecvf.com/content_CVPR_2019/html/Zellers_From_Recognition_to_Cognition_Visual_Commonsense_Reasoning_CVPR_2019_paper.html
AUTHORS: Rowan Zellers, Yonatan Bisk, Ali Farhadi, Yejin Choi
HIGHLIGHT: To move towards cognition-level understanding, we present a new reasoning engine, Recognition to Cognition Networks (R2C), that models the necessary layered inferences for grounding, contextualization, and reasoning.

TITLE: The Regretful Agent: Heuristic-Aided Navigation Through Progress Estimation
http://openaccess.thecvf.com/content_CVPR_2019/html/Ma_The_Regretful_Agent_Heuristic-Aided_Navigation_Through_Progress_Estimation_CVPR_2019_paper.html
AUTHORS: Chih-Yao Ma, Zuxuan Wu, Ghassan AlRegib, Caiming Xiong, Zsolt Kira
HIGHLIGHT: In this paper, inspired by the intuition of viewing the problem as search on a navigation graph, we propose to use a progress monitor developed in prior work as a learnable heuristic for search.

TITLE: Tactical Rewind: Self-Correction via Backtracking in Vision-And-Language Navigation
http://openaccess.thecvf.com/content_CVPR_2019/html/Ke_Tactical_Rewind_Self-Correction_via_Backtracking_in_Vision-And-Language_Navigation_CVPR_2019_paper.html
AUTHORS: Liyiming Ke, Xiujun Li, Yonatan Bisk, Ari Holtzman, Zhe Gan, Jingjing Liu, Jianfeng Gao, Yejin Choi, Siddhartha Srinivasa
HIGHLIGHT: We present the Frontier Aware Search with backTracking (FAST) Navigator, a general framework for action decoding, that achieves state-of-the-art results on the 2018 Room-to-Room (R2R) Vision-and-Language navigation challenge.

TITLE: Learning to Learn How to Learn: Self-Adaptive Visual Navigation Using Meta-Learning
http://openaccess.thecvf.com/content_CVPR_2019/html/Wortsman_Learning_to_Learn_How_to_Learn_Self-Adaptive_Visual_Navigation_Using_CVPR_2019_paper.html
AUTHORS: Mitchell Wortsman, Kiana Ehsani, Mohammad Rastegari, Ali Farhadi, Roozbeh Mottaghi
HIGHLIGHT: In this paper we study the problem of learning to learn at both training and test time in the context of visual navigation.

TITLE: Show, Control and Tell: A Framework for Generating Controllable and Grounded Captions
http://openaccess.thecvf.com/content_CVPR_2019/html/Cornia_Show_Control_and_Tell_A_Framework_for_Generating_Controllable_and_CVPR_2019_paper.html
AUTHORS: Marcella Cornia, Lorenzo Baraldi, Rita Cucchiara
HIGHLIGHT: In this paper, we introduce a novel framework for image captioning which can generate diverse descriptions by allowing both grounding and controllability.

TITLE: Towards VQA Models That Can Read
http://openaccess.thecvf.com/content_CVPR_2019/html/Singh_Towards_VQA_Models_That_Can_Read_CVPR_2019_paper.html
AUTHORS: Amanpreet Singh, Vivek Natarajan, Meet Shah, Yu Jiang, Xinlei Chen, Dhruv Batra, Devi Parikh, Marcus Rohrbach
HIGHLIGHT: Studies have shown that a dominant class of questions asked by visually impaired users on images of their surroundings involves reading text in the image.

TITLE: Object-Aware Aggregation With Bidirectional Temporal Graph for Video Captioning
http://openaccess.thecvf.com/content_CVPR_2019/html/Zhang_Object-Aware_Aggregation_With_Bidirectional_Temporal_Graph_for_Video_Captioning_CVPR_2019_paper.html
AUTHORS: Junchao Zhang, Yuxin Peng
HIGHLIGHT: In this paper, we propose a new video captioning approach based on object-aware aggregation with bidirectional temporal graph (OA-BTG), which captures detailed temporal dynamics for salient objects in video, and learns discriminative spatio-temporal representations by performing object-aware local feature aggregation on detected object regions.

TITLE: Progressive Attention Memory Network for Movie Story Question Answering
http://openaccess.thecvf.com/content_CVPR_2019/html/Kim_Progressive_Attention_Memory_Network_for_Movie_Story_Question_Answering_CVPR_2019_paper.html
AUTHORS: Junyeong Kim, Minuk Ma, Kyungsu Kim, Sungjin Kim, Chang D. Yoo

HIGHLIGHT: This paper proposes the progressive attention memory network (PAMN) for movie story question answering (QA).

TITLE: Memory-Attended Recurrent Network for Video Captioning
http://openaccess.thecvf.com/content_CVPR_2019/html/Pei_Memory-Attended_Recurrent_Network_for_Video_Captioning_CVPR_2019_paper.html
AUTHORS: Wenjie Pei, Jiyuan Zhang, Xiangrong Wang, Lei Ke, Xiaoyong Shen, Yu-Wing Tai
HIGHLIGHT: To tackle this limitation, we propose the Memory-Attended Recurrent Network (MARN) for video captioning, in which a memory structure is designed to explore the full-spectrum correspondence between a word and its various similar visual contexts across videos in training data.

TITLE: Visual Query Answering by Entity-Attribute Graph Matching and Reasoning
http://openaccess.thecvf.com/content_CVPR_2019/html/Xiong_Visual_Query_Answering_by_Entity-Attribute_Graph_Matching_and_Reasoning_CVPR_2019_paper.html
AUTHORS: Peixi Xiong, Huayi Zhan, Xin Wang, Baivab Sinha, Ying Wu
HIGHLIGHT: This paper proposes a novel method to address the VQA problem.

TITLE: Look Back and Predict Forward in Image Captioning
http://openaccess.thecvf.com/content_CVPR_2019/html/Qin_Look_Back_and_Predict_Forward_in_Image_Captioning_CVPR_2019_paper.html
AUTHORS: Yu Qin, Jiajun Du, Yonghua Zhang, Hongtao Lu
HIGHLIGHT: We propose Look Back (LB) method to embed visual information from the past and Predict Forward (PF) approach to look into future.

TITLE: Explainable and Explicit Visual Reasoning Over Scene Graphs
http://openaccess.thecvf.com/content_CVPR_2019/html/Shi_Explainable_and_Explicit_Visual_Reasoning_Over_Scene_Graphs_CVPR_2019_paper.html
AUTHORS: Jiaxin Shi, Hanwang Zhang, Juanzi Li
HIGHLIGHT: We aim to dismantle the prevalent black-box neural architectures used in complex visual reasoning tasks, into the proposed eXplainable and eXplicit Neural Modules (XNMs), which advance beyond existing neural module networks towards using scene graphs --- objects as nodes and the pairwise relationships as edges --- for explainable and explicit reasoning with structured knowledge.

TITLE: Transfer Learning via Unsupervised Task Discovery for Visual Question Answering
http://openaccess.thecvf.com/content_CVPR_2019/html/Noh_Transfer_Learning_via_Unsupervised_Task_Discovery_for_Visual_Question_Answering_CVPR_2019_paper.html
AUTHORS: Hyeonwoo Noh, Taehoon Kim, Jonghwan Mun, Bohyung Han
HIGHLIGHT: We tackle this problem in two steps: 1) learning a task conditional visual classifier, which is capable of solving diverse question-specific visual recognition tasks, based on unsupervised task discovery and 2) transferring the task conditional visual classifier to visual question answering models.

TITLE: Intention Oriented Image Captions With Guiding Objects
http://openaccess.thecvf.com/content_CVPR_2019/html/Zheng_Intention_Oriented_Image_Captions_With_Guiding_Objects_CVPR_2019_paper.html
AUTHORS: Yue Zheng, Yali Li, Shengjin Wang
HIGHLIGHT: In this paper, we propose a novel approach for generating image captions with guiding objects (CGO).

TITLE: Video Relationship Reasoning Using Gated Spatio-Temporal Energy Graph
http://openaccess.thecvf.com/content_CVPR_2019/html/Tsai_Video_Relationship_Reasoning_Using_Gated_Spatio-Temporal_Energy_Graph_CVPR_2019_paper.html
AUTHORS: Yao-Hung Hubert Tsai, Santosh Divvala, Louis-Philippe Morency, Ruslan Salakhutdinov, Ali Farhadi
HIGHLIGHT: In this paper, we construct a Conditional Random Field on a fully-connected spatio-temporal graph that exploits the statistical dependency between relational entities spatially and temporally.

TITLE: Image-Question-Answer Synergistic Network for Visual Dialog
http://openaccess.thecvf.com/content_CVPR_2019/html/Guo_Image-Question-Answer_Synergistic_Network_for_Visual_Dialog_CVPR_2019_paper.html
AUTHORS: Dalu Guo, Chang Xu, Dacheng Tao
HIGHLIGHT: In this paper, we devise a novel image-question-answer synergistic network to value the role of the answer for precise visual dialog.

TITLE: Not All Frames Are Equal: Weakly-Supervised Video Grounding With Contextual Similarity and Visual Clustering Losses
http://openaccess.thecvf.com/content_CVPR_2019/html/Shi_Not_All_Frames_Are_Equal_Weakly-Supervised_Video_Grounding_With_Contextual_CVPR_2019_paper.html
AUTHORS: Jing Shi, Jia Xu, Boqing Gong, Chenliang Xu
HIGHLIGHT: In this work, we address these issues by extending frame-level MIL with a false positive frame-bag constraint and modeling the visual feature consistency in the video.

TITLE: Inverse Cooking: Recipe Generation From Food Images
http://openaccess.thecvf.com/content_CVPR_2019/html/Salvador_Inverse_Cooking_Recipe_Generation_From_Food_Images_CVPR_2019_paper.html
AUTHORS: Amaia Salvador, Michal Drozdzal, Xavier Giro-i-Nieto, Adriana Romero
HIGHLIGHT: Therefore, in this paper we introduce an inverse cooking system that recreates cooking recipes given food images.

TITLE: Adversarial Semantic Alignment for Improved Image Captions
http://openaccess.thecvf.com/content_CVPR_2019/html/Dognin_Adversarial_Semantic_Alignment_for_Improved_Image_Captions_CVPR_2019_paper.html
AUTHORS: Pierre Dognin, Igor Melnyk, Youssef Mroueh, Jerret Ross, Tom Sercu
HIGHLIGHT: In this paper, we study image captioning as a conditional GAN training, proposing both a context-aware LSTM captioner and co-attentive discriminator, which enforces semantic alignment between images and captions.

TITLE: Answer Them All! Toward Universal Visual Question Answering Models
http://openaccess.thecvf.com/content_CVPR_2019/html/Shrestha_Answer_Them_All_Toward_Universal_Visual_Question_Answering_Models_CVPR_2019_paper.html
AUTHORS: Robik Shrestha, Kushal Kafle, Christopher Kanan
HIGHLIGHT: To address this problem, we propose a new VQA algorithm that rivals or exceeds the state-of-the-art for both domains.

TITLE: Unsupervised Multi-Modal Neural Machine Translation
http://openaccess.thecvf.com/content_CVPR_2019/html/Su_Unsupervised_Multi-Modal_Neural_Machine_Translation_CVPR_2019_paper.html
AUTHORS: Yuanhang Su, Kai Fan, Nguyen Bach, C.-C. Jay Kuo, Fei Huang
HIGHLIGHT: We propose an unsupervised multi-modal machine translation (UMNMT) framework based on the language translation cycle consistency loss conditional on the image, targeting to learn the bidirectional multi-modal translation simultaneously.

TITLE: Multi-Task Learning of Hierarchical Vision-Language Representation
http://openaccess.thecvf.com/content_CVPR_2019/html/Nguyen_Multi-Task_Learning_of_Hierarchical_Vision-Language_Representation_CVPR_2019_paper.html
AUTHORS: Duy-Kien Nguyen, Takayuki Okatani
HIGHLIGHT: We propose a multi-task learning approach that enables to learn vision-language representation that is shared by many tasks from their diverse datasets.

TITLE: Cross-Modal Self-Attention Network for Referring Image Segmentation
http://openaccess.thecvf.com/content_CVPR_2019/html/Ye_Cross-Modal_Self-Attention_Network_for_Referring_Image_Segmentation_CVPR_2019_paper.html
AUTHORS: Linwei Ye, Mrigank Rochan, Zhi Liu, Yang Wang
HIGHLIGHT: In this paper, we propose a cross-modal self-attention (CMSA) module that effectively captures the long-range dependencies between linguistic and visual features.

TITLE: Good News, Everyone! Context Driven Entity-Aware Captioning for News Images
http://openaccess.thecvf.com/content_CVPR_2019/html/Biten_Good_News_Everyone_Context_Driven_Entity-Aware_Captioning_for_News_Images_CVPR_2019_paper.html
AUTHORS: Ali Furkan Biten, Lluis Gomez, Marcal Rusinol, Dimosthenis Karatzas
HIGHLIGHT: In this work, we aim to take a step closer to producing captions that offer a plausible interpretation of the scene, by integrating such contextual information into the captioning pipeline.

TITLE: Multi-Level Multimodal Common Semantic Space for Image-Phrase Grounding
http://openaccess.thecvf.com/content_CVPR_2019/html/Akbari_Multi-Level_Multimodal_Common_Semantic_Space_for_Image-Phrase_Grounding_CVPR_2019_paper.html
AUTHORS: Hassan Akbari, Svebor Karaman, Surabhi Bhargava, Brian Chen, Carl Vondrick, Shih-Fu Chang
HIGHLIGHT: We address the problem of phrase grounding by learning a multi-level common semantic space shared by the textual and visual modalities.

TITLE: Spatio-Temporal Dynamics and Semantic Attribute Enriched Visual Encoding for Video Captioning
http://openaccess.thecvf.com/content_CVPR_2019/html/Aafaq_Spatio-Temporal_Dynamics_and_Semantic_Attribute_Enriched_Visual_Encoding_for_Video_CVPR_2019_paper.html
AUTHORS: Nayyer Aafaq, Naveed Akhtar, Wei Liu, Syed Zulqarnain Gilani, Ajmal Mian
HIGHLIGHT: These methods mainly focus on tailoring sequence learning through RNNs for better caption generation, whereas off-the-shelf visual features are borrowed from CNNs.

TITLE: Pointing Novel Objects in Image Captioning
http://openaccess.thecvf.com/content_CVPR_2019/html/Li_Pointing_Novel_Objects_in_Image_Captioning_CVPR_2019_paper.html
AUTHORS: Yehao Li, Ting Yao, Yingwei Pan, Hongyang Chao, Tao Mei
HIGHLIGHT: In this paper, we propose to address the problem by augmenting standard deep captioning architectures with object learners.

TITLE: Informative Object Annotations: Tell Me Something I Don't Know
http://openaccess.thecvf.com/content_CVPR_2019/html/Bracha_Informative_Object_Annotations_Tell_Me_Something_I_Dont_Know_CVPR_2019_paper.html
AUTHORS: Lior Bracha, Gal Chechik
HIGHLIGHT: Motivated by cognitive theories of categorization and communication, we present a new unsupervised approach to model this prior knowledge and quantify the informativeness of a description.

TITLE: Engaging Image Captioning via Personality
http://openaccess.thecvf.com/content_CVPR_2019/html/Shuster_Engaging_Image_Captioning_via_Personality_CVPR_2019_paper.html
AUTHORS: Kurt Shuster, Samuel Humeau, Hexiang Hu, Antoine Bordes, Jason Weston
HIGHLIGHT: We build models that combine existing work from (i) sentence representations [36] with Transformers trained on 1.7 billion dialogue examples; and (ii) image representations [32] with ResNets trained on 3.5 billion social media images.

TITLE: Vision-Based Navigation With Language-Based Assistance via Imitation Learning With Indirect Intervention
http://openaccess.thecvf.com/content_CVPR_2019/html/Nguyen_Vision-Based_Navigation_With_Language-Based_Assistance_via_Imitation_Learning_With_Indirect_CVPR_2019_paper.html
AUTHORS: Khanh Nguyen, Debadeepta Dey, Chris Brockett, Bill Dolan
HIGHLIGHT: To model language-based assistance, we develop a general framework termed Imitation Learning with Indirect Intervention (I3L), and propose a solution that is effective on the VNLA task.

TITLE: TOUCHDOWN: Natural Language Navigation and Spatial Reasoning in Visual Street Environments
http://openaccess.thecvf.com/content_CVPR_2019/html/Chen_TOUCHDOWN_Natural_Language_Navigation_and_Spatial_Reasoning_in_Visual_Street_CVPR_2019_paper.html
AUTHORS: Howard Chen, Alane Suhr, Dipendra Misra, Noah Snavely, Yoav Artzi
HIGHLIGHT: We study the problem of jointly reasoning about language and vision through a navigation and spatial reasoning task.

TITLE: A Simple Baseline for Audio-Visual Scene-Aware Dialog
http://openaccess.thecvf.com/content_CVPR_2019/html/Schwartz_A_Simple_Baseline_for_Audio-Visual_Scene-Aware_Dialog_CVPR_2019_paper.html
AUTHORS: Idan Schwartz, Alexander G. Schwing, Tamir Hazan
HIGHLIGHT: Therefore, in this paper, we provide and carefully analyze a simple baseline for audio-visual scene-aware dialog which is trained end-to-end.