Automatic image annotation assigns relevant textual tags to the images. We propose a mathematical framework based on Non-negative Matrix Factorization to perform automatic image annotation. Our proposed query-specific approach naturally solves the feature fusion and handles the challenge of rare tags.

Unlike tags whose descriptive power scales linearly, attributes model an exponential number of semantic classes. We hypothesize that integrating human semantic parsing improves person-related visual attribute prediction. In this regard, we propose Semantic Segmentation-based Gating (SSG) and Pooling (SSP). In SSG, we create multiple copies where each preserves the activations within a single semantic region and suppresses otherwise. This mechanism prevents semantically inconsistent regions to be mixed. SSP behaves similarly but at the classifier layer. To tackle the heavy memory utilization of SSP and SSG, Symbiotic Augmentation is proposed, where we learn to generate only one mask per activation channel.

Given Big Data nature of Selfies, it is impossible to analyze them manually. Next, we use both textual tags and visual attributes to analyze Selfies. We collect the first Selfie dataset with more than 46K images and annotate it with 36 visual attributes covering characteristics such as gender, age, race, and hairstyle. We provide attribute prediction of Selfies and by training regularized SVR for the normalized view counts, assess the impact of different visual concepts and various Instagram filters on the popularity of Selfies.

Almost all today’s deep convolutional neural networks use Batch Normalization (BN), yet its characteristics are not sufficiently studied in the literature. We conclude this dissertation by revisiting BN in light of Fisher vector of Gaussian density function. We offer theoretical and experimental support that disentangling modes of variation in the underlying distribution of layer outputs not only effectively accelerates training of different batch-normalized architectures including Inception-V3, DenseNet, and DCGAN, but also achieves better generalization error.