RESEARCH STATEMENT

Ze Wang

My Ph.D. research focuses on computer vision models that efficiently learn and adapt with minimal supervision. Motivated by this objective, my research has progressed in three main directions: (1) vision model adaptation in low-rank filter subspace; (2) few-shot adaptive vision models; and (3) representation learning with generative models. Below, I will describe the current progress of my research in each of the three directions. And I will conclude by discussing my ongoing goals and future interests.

Vision Model Adaptation in Low-rank Filter Subspace

Filters in convolutional neural networks are usually of high dimension, which makes network adaptation both challenging and costly. Inspired by the fact that a convolutional filter can be well represented as a linear combination of a small set of filter atoms [1], I developed convolutional vision models that adapt the filters in low-rank filter subspace, which achieves efficient model adaptation to novel domains with weak supervision [2]; generative models that perform conditional sampling with strong diversity [3] and smooth appearance control [4]; continual learning by swapping filter subspace [5]; and vision models that apply *per-pixel dynamic convolutional filters* with respect to the features at each *spatial position* [6]. Combining with a joint subspace view of convolutional neural networks [7], the per-pixel adaptive convolution filters potentially allow neural networks to grow with arbitrary depth and layer-specific dynamic parameters.

Few-shot Adaptive Vision Models

The real-world performance of learning models can be largely determined by how well prior knowledge acquired during training is utilized to efficiently adapt models to novel concepts using a few examples. With dense inducing variables specifying a shared Gaussian process prior over prediction functions of various tasks, and efficient gradient-based adaptation to the inducing variables only, I demonstrated in [8] that Gaussian processes with deep neural networks are strong few-shot learners with well-captured uncertainty. And with robust geometry regularization to the learned feature space to promote both intra-class low-rankness and inter-class orthogonality, a few-shot classification network can demonstrate enhanced generalization to novel classes with insufficient annotations and effectively exploit unlabeled data for task-specific adaptation [9].

Representation Learning with Generative Models

The potential for exploring useful and transferable knowledge from massive unlabeled vision data and multi-modal data stimulates my recent efforts on unsupervised representation learning with generative models. I developed binary latent diffusion models [10] with the diffusion and the corresponding denoising processes for multivariate Bernoulli distribution, which are able to effectively model the image distribution in a learned binary latent space. In addition, the proposed binary latent diffusion exhibits exceptional sampling speed, with high-quality images generated in as few as eight denoising steps, and high-resolution images generated in one-shot. In order to learn transferable knowledge from unlabeled data, I extended probabilistic modeling of data distribution with energy-based models (EBMs) to learning general and generalizable image representations [11]. In the proposed framework, any vision model can be pretrained by learning to identify objects given partial observations of images and restore the missing contents based on the knowledge residing in the model parameters using gradient-based updating to the input pixel values. Using a novel adaptive sparse coding layer, I further demonstrated in [12] that adaptive EBMs are capable of few-shot fast-adaptive anomaly detection for novel objects with only a few normal samples and no parameter tuning.

Ongoing Work and Interests

My ongoing research focuses primarily on further expanding the capability of low-rank model adaptation, and adaptive and transferable representation learning using generative models with minimal annotations. I am particularly interested in investigating the following areas:

- *Efficient model adaptation to novel concepts and tasks.* Large pretrained large models have demonstrated unprecedented capabilities. However, their downstream application typically focuses on specific tasks or knowledge that are not always adequately covered by pretraining. Improving transferability can be achieved by developing efficient post-training adaptation methods that are compatible with various pretrained models, or by directly incorporating and learning adaptive model components in the pretraining stage [12].
- *Deep models with adaptive computation.* Most modern deep neural networks apply the same computation to all samples. With the model size expanding rapidly to accommodate the difficult samples, massive computation can be wasted on simple ones, preventing the widespread deployment of learning systems in practical applications. Therefore, I am interested in exploring vision models with dynamic architectures and depth, which can dynamically budget the computation cost on each sample based on sample-wise difficulties and prediction confidence. Importantly, this must be accomplished without additional annotations during training.
- *Versatile learning models.* Modern deep neural networks use parameters as the only source of knowledge, necessitating a growing number of parameters to compress and store all training knowledge. My long-term research interest is in developing versatile learning models that are able to seamlessly handle data from various modalities, and exploit knowledge from diverse sources, e.g., external databases and the Internet, in an adaptive and associative manner.

I envision my research empowering adaptive AI systems that efficiently master novel concepts and tasks by leveraging knowledge extracted from massive data with minimal supervision both during and after training.

References

[1] Qiang Qiu, Xiuyuan Cheng, Guillermo Sapiro, et al. DCFNet: Deep neural network with decomposed convolutional filters. In *ICML*, 2018.

[2] **Ze Wang**, Xiuyuan Cheng, Guillermo Sapiro, and Qiang Qiu. A dictionary approach to domain-invariant learning in deep networks. *NeurIPS*, 2020.

[3] **Ze Wang**, Xiuyuan Cheng, Guillermo Sapiro, and Qiang Qiu. Stochastic conditional generative networks with basis decomposition. *ICLR*, 2020.

[4] **Ze Wang**, Seunghyun Hwang, Zichen Miao, and Qiang Qiu. Image generation using continuous filter atoms. *NeurIPS*, 2021.

[5] Zichen Miao, **Ze Wang**, Wei Chen, and Qiang Qiu. Continual learning with filter atom swapping. In *ICLR*, 2022.

[6] **Ze Wang**, Zichen Miao, Jun Hu, and Qiang Qiu. Adaptive convolutions with per-pixel dynamic filter atom. In *ICCV*, 2021.

[7] **Ze Wang**, Xiuyuan Cheng, Guillermo Sapiro, and Qiang Qiu. ACDC: Weight sharing in atom-coefficient decomposed convolution. Under review.

[8] **Ze Wang**, Zichen Miao, Xiantong Zhen, and Qiang Qiu. Learning to learn dense gaussian processes for few-shot learning. *NeurIPS*, 2021.

[9] **Ze Wang**, Yue Lu, and Qiang Qiu. Meta-ole: Meta-learned orthogonal low-rank embedding. In *WACV*, 2023.

[10] **Ze Wang**, Jiang Wang, Zicheng Liu, and Qiang Qiu. Binary latent diffusion. Under review.

[11] **Ze Wang**, Jiang Wang, Zicheng Liu, and Qiang Qiu. Energy-inspired self-supervised pretraining for vision models. In *ICML*, 2022, Pretraining Workshop. Full version under review.

[12] **Ze Wang**, Yipin Zhou, Rui Wang, Tsung-Yu Lin, Ashish Shah, and Ser-Nam Lim. Few-shot fast-adaptive anomaly detection. In *NeurIPS*, 2022.