Subhransu Maji

Research Statement

I am a computer scientist who conducts research in the areas of *computer vision* and *machine learning*. My goal as a researcher is to make fundamental contributions towards building AI systems with rich visual reasoning capabilities. My research focuses on *architectures* for visual recognition tasks [1–5, 7–14, 19, 25, 26, 29, 35, 42–45], as well as techniques to improve their *robustness*, *efficiency*, *generalization* and *interpretability* [15–17, 20–24, 30–34, 36–39, 41, 46–53]. My research is also *interdisciplinary*. I collaborate with computer scientists and ecologists to analyze bird migration from radar imagery; with astronomers to uncover scientific insights from images of galaxies; as well as domain experts to develop applications in graphics, health care, material design, and sustainability. I am a long-term organizer of the *Fine-Grained Visual Categorization* workshop which aims to foster collaborations between researchers in computer vision, ecology, biology and industry, and to understand the role of computer vision in other domains.

My research is motivated not only by the numerous applications of computer vision, but also by the possibility that fundamental advances in analyzing visual data can provide insights into various social, cultural, and natural factors that profoundly affect our lives. For example, data collected from personal devices, social media, autonomous platforms and medical devices can support novel applications in e-commerce, robotics, manufacturing, and health care. Similarly, data collected by satellites, weather radars, telescopes and other sensor networks provide an unprecedented opportunity to *understand* planet-scale phenomena such as the impact of climate change on biodiversity and shed light into processes that govern the physical world. I argue that a primary barrier to realizing the full potential of these data resources is *computational*. In particular, current AI systems lack the ability to perceive detailed semantic information, to work across modalities such a language and vision, interact with humans to solve problems, and to solve novel tasks with limited supervision.

Application areas. My research addresses these challenges in the context of the following application areas, as illustrated in Fig. 1–4:

- *Fine-grained and detailed visual recognition.* Detailed understanding is necessary to enable richer applications of computer vision across domains ranging from e-commerce, to biology and ecology. We have developed architectures for fine-grained categorization tasks such as identifying the species of a bird, or the model of a car (e.g., [25, 26]). We have also developed methods for understanding *texture*, which can be used to estimate material properties of surfaces or clothing styles (e.g., [8, 9]). In my PhD thesis, I developed part-based models for estimating attributes of people such as their pose, activities, and clothing styles in photographs (e.g., [3–5, 35]).
- *3D shape understanding and generation.* We have developed architectures for 3D shape perception tasks such as classification and semantic segmentation (e.g., [44, 45]), for estimating shapes from a single image, image collections, or even sketches (e.g., [11, 13, 29]), as well as methods for decomposing 3D data into primitives for easier editing and manipulation. These methods can enable applications in robotics and autonomous driving where 3D understanding is necessary for navigation, grasping, and other tasks.
- *Applications in computer graphics, ecology, astronomy and chemistry.* With collaborators, we have developed novel applications of computer vision to problems in *computer graphics*, such as animation and editing textures using language; in *ecology*, such as classifying animals in camera-trap images and recognizing biological phenomena in radar imagery; in *astronomy*, such as for classifying star clusters in high resolution images of galaxies (e.g., [28, 41, 49, 53]); with *chemists* to develop ways to represent and design catalysts for chemical separation.

Research themes. In the context of the aforementioned applications my research follows two themes:

- *Learning and recognition with humans in-the-loop.* We have developed annotation tasks and learning algorithms to discover semantic parts and attributes associated with a category and to train interpretable visual representations [30, 38]; user-interfaces to efficiently annotate detailed labels [36]; and techniques for incorporating human knowledge during inference making it possible to deploy imperfect recognition systems for solving complex tasks [51, 52].
- *Statistical and computational efficiency in learning.* My PhD thesis contributed to improving the efficiency of recognition systems (e.g., [33, 37]). Recently, we have analyzed the role of image representations and architectures for 3D shape understanding, developed algorithms for few-shot learning and for learning with weak supervision, and developed a framework for measuring similarity between tasks for meta-learning (e.g., [20, 46–48]).

Research collaboration, impact, and funding. Within computer science, I am a member of the computer vision and machine learning (broadly AI) communities. My research strategy is to *publish* fundamental and application-driven results in computer vision (CV), graphics and machine learning (ML) venues, and *collaborate* with domain experts to publish novel scientific findings in ecology and astronomy. I have organized the last *eight* Fine-Grained Visual Categorization workshops (FGVC^3— FGVC^9), and contributed several datasets and open-source software to the AI community. Since joining UMass, I have published 38 articles in highly selective CV, ML and AI conferences, including 27 at CVPR, ICCV, ECCV, KDD, SIGGRAPH, and AAAI, nine journal articles at selective venues, and two book chapters. My publications have been cited *16196 times* (*h-index 41*, *i10-index 67*) according to Google scholar as of March 8, 2022. Papers that I have co-authored have received the *best paper honorable mention* at CVPR 2018 [44], and *best student paper* at WACV 2015 [52]. My work has been supported by *six awards* from the National Science Foundation, Climate Change AI foundation, with two as sole PI and four as Co-PI, and gifts from Facebook, Adobe, and NVIDIA; the UMass portions of these awards total 3.2 million US dollars. On the education side, I have supervised nine PhD students (five ongoing and four have graduated).

Below I describe my research contributions and future work organized into three thrusts: *architectures for visual recognition*, techniques for *knowledge transfer and learning*, and applications of computer vision in *ecology, astronomy, and other areas*.

1 Architectures for Visual Recognition

Advances in computer vision, in part due to the "deep learning" revolution, offer a real possibility of *universal* and *detailed* visual perception. A challenge is that current techniques are reliant on significant manual supervision, limiting their scalability to tasks and domains. My research develops architectures with better statistical and computational efficiency to address these challenges. My research also develops novel architectures for analyzing and generating modalities such as texture and 3D data. I will highlight these in the context of various problems that I have tackled recently.

1.1 Texture Understanding and Fine-Grained Categorization

Texture is indicative of an object's shape, material properties, and also useful for fine-grained categorization tasks, such as distinguishing animal species as illustrated in Fig. 1(a). The ability to perform these tasks increases the value of data collected through a variety of sensors, e.g., to estimate distribution of species and inform conservation efforts, as well as enable applications in robotics and computer graphics. In a series of work that started as a collaboration with Mircea Cimpoi, Andrea Vedaldi and Iasonas Kokkinos, and continued by my students, we have developed a better understanding of how texture properties can be analyzed with deep networks, and in the process proposed several innovative architectures for texture and

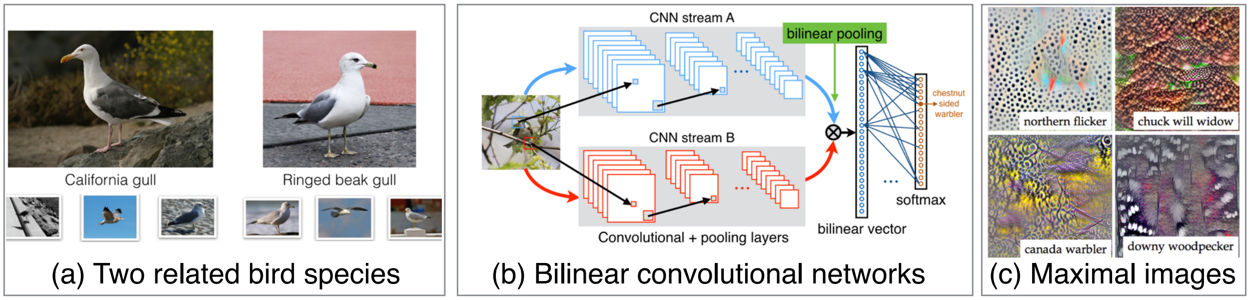

Figure 1: **Fine-grained visual categorization.** **(a)** Two visually similar species of Gulls. Below each are examples of each species; notice the inter- and intra-class variability. **(b)** Bilinear CNNs compute outer product interactions between features from two streams and was one of the early methods proposed by our group that achieved high performance on this task. **(c)** Visualizing "maximal images" of a category according to a bilinear CNN shows the discriminative "texture" indicative of that category.

fine-grained categorization. I will summarize the main contributions below:

- At CVPR 2014, we presented the **describable texture dataset** aimed at describing textures "in the wild" using natural language [8]. The dataset served as a tesbed for analyzing the effectiveness of deep networks trained on ImageNet for the classical problem of texture and material recognition.
- At CVPR 2015 [10] and in a subsequent IJCV journal [9], we observed that though the features extracted from layers of a deep convolution neural network (CNN) offered excellent generalization on many computer vision datasets, they were not as effective for texture recognition compared to classical approaches. We showed that instead, combining features from intermediate layers of a deep network with *encoding* and *orderless pooling* techniques developed in the classical texture analysis literature led to better performance. In particular, we proposed **Fisher vector CNNs** that outperformed *both* classical representations and standard CNN architectures across a wide variety of texture datasets.

We then studied the role of texture for *fine-grained categorization*. This task is challenging because the subtle differences between categories are confounded by factors such as pose, viewpoint, and occlusion (Fig. 1a). The dominant approach at the time was to localize a set of parts and extract a pose-normalized appearance. While these methods were more accurate, training part detectors required significant effort as they relied on part annotations.

- At ICCV 2015, we proposed **bilinear CNN** [25, 26], a deep architecture that offered the effectiveness of part-based models but did *not* require part annotations. The key idea was a decomposition of the representation as an outer product (see Fig. 1b) designed to capture discriminative part-feature interactions. We also showed that the model is related to bag-of-words representations, but unlike them could be trained in an end-to-end manner. The model surpassed the state-of-the-art on several fine-grained recognition benchmarks, even outperforming several that relied on part annotations. The work was influential in the design of several architectures for fine-grained classification, and the idea of combining of information from multiple streams through outer product interactions has found its use in tasks such as visual question answering and activity recognition.
- At CVPR 2016, we developed a technique [22] to analyze what these models learn by visualizing maximal images for each category as shown in Fig. 1c. Recently, we applied this technique to visualize and describe various fine-grained datasets from the **FGVC6 workshop challenges** [27], shedding some light on the discriminative visual features predictive of the category.
- At BMVC 2017, we proposed the **improved bilinear CNN** [23] that increased the effectiveness of bilinear pooling by using spectral normalization. Furthermore, we showed that the procedure could be

efficiently implemented on the GPU using iterative methods. At ECCV 2018, we presented **second-order democratic pooling** [24] which is simpler and more efficient. These models remain effective, with a team winning the iNaturalist18 challenge by combining improved bilinear CNNs with deeper networks. It was also a component of some of the top-performing methods at the FGVC6 challenges.

1.2 3D Shape Recognition and Generation

There is a growing need to analyze and generate 3D shape data for applications in computer graphics, robotics and autonomous driving. However, existing techniques for processing 3D data are lacking in comparison to those for image data. In the last few years, together with colleagues at UMass, in particular Erik Learned-Miller, Rui Wang and Evangelos Kalogerakis, and several graduate students, in particular Matheus Gadelha, Hang Su and Jong-Chyi Su, we have advanced techniques for 3D shape analysis and synthesis as illustrated in Fig. 2. I will summarize the main contributions below.

- At ICCV 2015, we presented a deep architecture for 3D shape recognition called **multi-view CNN** [45] (Fig. 2a). The technique aggregates view-based representations from multiple views to obtain a global 3D shape representation. The architecture is end-to-end trainable and benefits from transferable representations learned on large-scale image datasets. At the time of publication, this model achieved state-of-the-art performance on standard 3D shape classification and retrieval tasks (e.g., [42]). Although several new techniques have been proposed since, a recent survey from our group presented at an ECCV 2018 workshop [46] showed that multi-view architectures outperform these techniques when they are combined with the latest image classification networks.
- At CVPR 2017, in a work led by Evangelos Kalogerakis, we presented **Shape-PFCN** [19], an architecture for 3D shape segmentation. Segmentations from multiple views were projected back to the surface followed by global reasoning to obtain a consistent segmentation. The approach is simple and practical and has been effectively applied to other 3D semantic segmentation tasks.
- At 3DV 2017, we presented an approach for generating 3D shapes from sketches using a mutli-view depth and normal representation of the 3D shape, as illustrated in Fig. 2b. The key insight was to generate vast amounts of synthetic sketch data by applying the techniques developed in the computer graphics community to supervise deep networks. This was a novel application of deep learning for the classical task of understanding shape from sketches that pioneers like Jan J. Koenderink studied.
- At 3DV 2017, we presented **PRGAN** [11], an approach for inferring 3D shapes from shape collections that did not require any prior knowledge about the underlying shapes. This is a challenging problem related to non-rigid structure from motion. Our approach combined a 3D generative model with a differentiable renderer, and used adversarial networks to match the generated image distributions to the provided set of images. The method was able to infer coarse 3D shapes and viewpoints in a *completely unsupervised manner*. In a recent work [12] we extended the approach to incorporate weak supervision, such as viewpoint and part annotations, to improve the reconstructions.
- At CVPR 2018, we presented **SPLATNet** [44], an architecture that efficiently operates on point-cloud data using sparse high-dimensional filtering, as illustrated in Fig. 2c. The approach also allows one to effectively combine view-based and 3D-based reasoning for recognition tasks. Moreover the architecture can also be efficiently implemented on the GPU making it practical for real-time applications. The approach was awarded **best paper honorable mention** at CVPR 2018 (1/4 of 979 accepted papers).
- At ECCV 2018, we presented one of the first generative models for point-cloud data called **MRTNet** [13]. The idea was to parameterize the point-cloud data implicitly using a spatial partitioning tree and use the induced hierarchy to organize feed-forward operations in a generative model. The model can be used to encode point cloud shapes or generate 3D shapes from images as shown in Fig. 2(d-e). Our approach outperformed the existing state-of-the-art on standard image to shape generation benchmarks. The work improved our earlier generative model for point clouds called **PCAGAN** [14].

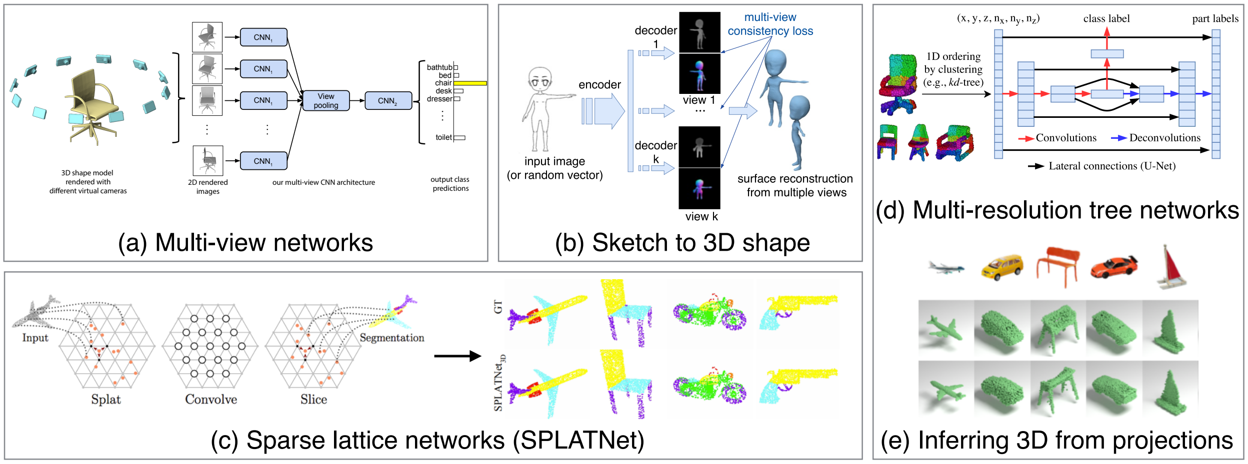

Figure 2: **3D shape understanding.** (a) Multi-view CNN for 3D shape recognition. **(b)** Generating 3D shapes from sketches. **(c)** SPLATNet architecture for shape segmentation. **(d-e)** Multiresolution tree networks (MRTNet) and its application to generate 3D shapes represented as point clouds.

2 Knowledge Transfer and Learning

Humans are able to recognize new concepts from a few examples while current AI systems cannot. My research addresses this limitation in two ways. First, I study how humans can efficiently collaborate with computer vision systems to train them and solve complex recognition tasks. Second, I study ways to improve the generalization and computational efficiency of recognition systems by analyzing the role of datasets, architectures and training algorithms, contributing to the growing area of research on meta-learning. I will highlight the main contributions next.

2.1 Efficiency of Learning and Inference

My PhD thesis proposed ways to improve the efficiency and accuracy of detection and classification systems:

- **Additive kernel SVMs** [32–34] that reduced the training and evaluation time for several commonly used kernel SVMs (e.g., intersection, chi-squared) in computer vision, making them practical for detection and other large-scale classification tasks. This work was influential in the development of several "pre deep-learning" computer vision systems, including the region-based pipeline for object detection that won the PASCAL VOC challenges 2009-11, a precursor to the modern R-CNN detectors.
- **Poselets** that are visually discriminative patterns and capture a configuration of parts of an object. In a series of work in collaboration with several people, in particular Lubomir Bourdev, we showed that poselets were effective for detecting objects and recognizing their attributes [3, 4, 35].
- **Biased normalized cuts** [39] that enabled interactive image segmentation from scribbles and clicks in the normalized-cuts framework.
- **Learnable Hough transforms** for combining part-based models for object detection [37].

Together with collaborators I have continued this line of research.

- In a project led by Tamir Hazan and in collaboration with several researchers, we developed a **Perturb and MAP** framework for drawing samples from probabilistic models by perturbing them and using efficient MAP inference algorithms. This enabled inference and learning of certain structured models [15–17]. At AISTATS 2014, we used this framework for accelerating annotation of structured data, such as polygon boundaries of objects in images, by jointly reasoning about annotation cost and uncertainty induced by a probabilistic model [36].

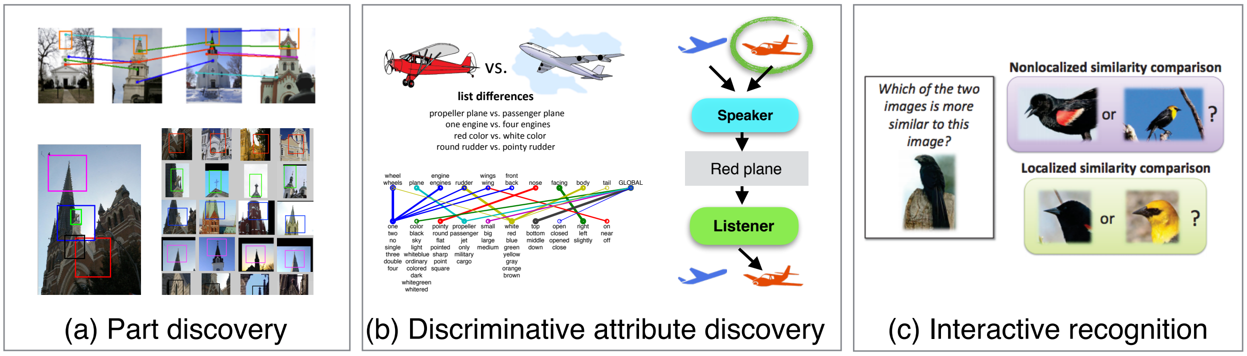

Figure 3: **Recognition with humans in-the-loop (a)** Discovering semantic parts from pairwise correspondence annotations. **(b)** Attribute discovery from visual differences and a speaker-listener model trained to refer to instances. **(c)** Localized similarity metrics for interactive recognition.

- At ECCV 2014, we proposed a technique to accelerate object detectors by training a model to predict the accuracy of the detector using features computed from the model itself, speeding the overall training by an order of magnitude [2].
- At CVPR 2019, we presented an approach for few-shot learning called **CVXOptNet** [20]. The idea was to learn representations that lead to good classifiers according to a *convex learner*. This resulted in consistent improvements over the commonly used nearest neighbor classifiers at a modest increase in computational cost on standard few-shot learning benchmarks.
- At CVPR 2019, we characterized the priors induced by a random deep network by analyzing it as a Gaussian process, and proposed ways to improve the "deep image prior" of Ulyanov et al. for image denoising and inpainting tasks [7].

2.2 Learning and Recognition with Humans in-the-Loop

Another line of research aims at simplifying how to collect annotations for training computer vision systems. We have developed techniques where semantic properties can be discovered by asking human annotators to perform simple tasks that don't require any category-specific instructions.

- At CVPR 2012, we showed that by asking annotators to **mark correspondences** between pairs of images one can discover semantic parts as seen in Fig. 3a. These parts provide an interpretable basis for detection and retrieval [38].
- At CVPR 2013, we showed that by asking annotators to **describe the differences** between instances one can discover a rich lexicon of discriminative attributes [30], as seen in Fig. 3b). These ideas were instrumental in the collection of the **airplanes in detail** dataset [50].
- At ICCV 2017, we showed that training computer vision systems to analyze and generate visual differences lead to better models for language and vision tasks [48] (see Fig. 3b).

Interpretable representations not only make subsequent debugging and visualization of the learned models easier for humans, it can serve as a basis for interactive recognition. For example:

- We developed techniques for recognition with **humans in-the-loop** (e.g. [51, 52]), where perceptually-aligned representations can be used to incorporate human feedback for fine-grained categorization tasks (Fig. 3c). The latter [52] was awarded a **best student paper** at WACV 2015.
- At BMVC 2017, we studied *how* annotations can be incorporated to improve learning. Typically, part labels or attributes are used in "pipelined" architecture (e.g., detect and then classify). However, simpler end-to-end trained models are more effective as more training data becomes available. We presented a

technique called **cross quality distillation** to use annotations as *loss functions* whose influence could be adjusted based on the amount of training data. Our work showed that off-the-shelf models trained with these loss functions match the accuracy of hand-designed pipelines [47].

- In collaboration with Evangelos Kalogerakis, we have developed interpretable generative models for computer graphics applications where it is often desirable to have an *editable representation*. At CVPR 2018, we presented **CSGNet** that decomposes 3D shapes into a computer program written in constructive solid geometry [43]. Similarly, in SIGGRAPH 2018, we presented **VISEMENet** to generate editable facial animation parameters controlled by speech [53].

Human in-the-loop recognition is also relevant for quickly building systems for solving novel tasks. In this context methods for active learning, semi-supervised learning, and interaction design become important to develop good workflows for data processing.

3 Applications

Ecology. I lead the computer vision efforts for *dark ecology* project — a significant ongoing research collaboration between the University of Massachusetts and the Cornell Lab of Ornithology. Our goal is to quantitatively measure biological information from the entire archive of US weather RADAR data and use the resulting data to unravel mysteries of bird migration. We have developed automatic techniques to separate weather from biology in radar data so we can accurately and automatically measure migration, and cloud workflows to scale our analyses to this massive data set (Fig. 4a). Recently, we passed a significant milestone: our model now identifies at least 95.9% of all biomass with a false discovery rate of 1.3% across a wide range of conditions [28], and we have analyzed 13 million scans using more than 10,000 hours of compute time to construct a 25-year data set of migration covering all 143 radar stations in the contiguous US at a frequency of once every 30 minutes. The data is the first of its kind for answering historical questions about the basic biology and conservation of bird migration systems. For example, we found that the phenology of nortornal bird migration has shifted at by $\approx$ 1.5 days / decade in the continental US [18]. Another ongoing project is aimed at detecting roosts [6] of birds and bats which have a distinctive signature on weather radar, seen as the ring-shaped structures in Fig. 4b. These phenomenon offer insights to the migratory behavior of aerial insectivores which are threathened by a number of factors. These projects are actively funded by two grants from the NSF, for which I am a co-PI.

Astronomy. I collaborate with Daniela Calzetti from the Astronomy department, to develop algorithms for search, classification, and shape measurement of young star clusters in high resolution images of galaxies. Our current techniques are able to achieve accuracy comparable to human experts [40] and we are evaluating how well these methods generalize to novel tasks. Our initial tests were performed on the two closest galaxies to our own Milky Way (M31 and M33), and then extended to M51 and NGC628, which are further away from us. These are well-studied galaxies for which high-fidelity catalogs already exist, and can be used for training and evaluation of automatic systems. We anticipate that data collected using these techniques will provide insights into the physics of star formation, and serve as pilot studies for building efficient workflows for processing future data that will arrive through the James Webb telescope.

Others collaborations. In collaboration with students and faculty members I developed applications of computer vision that include: a dataset of camera trap images from Costa Rica and algorithms for recognizing the presence and species of animals that appear in them [49]; algorithms for inferring 3D shape of rooftops from satellite imagery to estimate their solar power potential [21]; and using computer vision to train a model for predicting facial expressions from Electrooculography (EOG) based sensors embedded in wearable eye glasses [41]. An ongoing collaboration with the Chemistry department aims at developing representations of extended materials such as zeolites, for designing catalyts for chemical seperation.

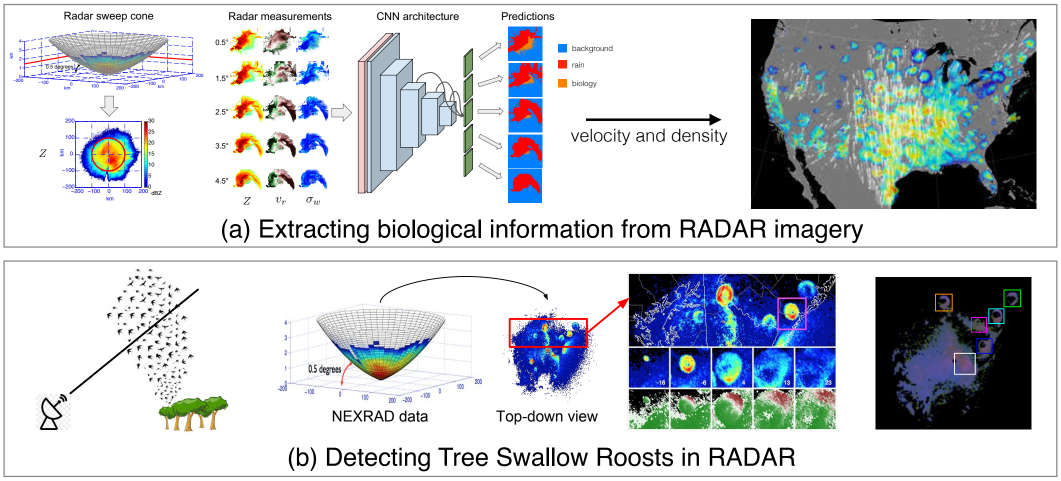

Figure 4: **The dark ecology project to understand bird migration from RADAR data in the US.**

4 Future work and conclusion

My research group is investigating several projects for manipulating, generating, and performing inference collaboratively between humans and machines. One direction is for computer graphics applications where raw data such as point-clouds are represented using primitives for easier manipulation and editing. Another project aims to use natural language to generate and describe regions in images for interactive editing. Another is to accelerate science where expert labels can be used to train ML systems to process vast amount of scientific data, or serve as surrogate models for experimental design.

My research group is also looking into language as a means of communication for solving complex tasks with humans and AI systems. Current models for integrating language and vision are one-directional, i.e., machines are trained to produce descriptions in natural language, or answer questions about an image or video. We would like to develop systems that can truly communicate, i.e., machines that can ask questions to experts and vice-versa to solve visual recognition tasks.

Another direction in our group aims to improve the generalization of recognition systems by analyzing the space of vision tasks. Recently, we developed a framework for efficiently computing vectorial representations of computer vision tasks [1]. The distance in the embedding space reflects task similarity and can be used for model recommendation and other meta-learning tasks. Understading and relating tasks and their solutions across diverse will enable us to build systems that can quickly learn and adapt to novel tasks.

My research group is also exploring ways to improve generalization of models when learning from a few examples by exploiting unlabeled data from different modalities; improving generative models of 3D shapes by analyzing inductive biases of deep architectures; developing models for analyzing spatio-temporal data such as video and audio; increasing the robustness of 3D shape inference from images in realistic settings where they could be occluded or appear in clutter; as well as methods for inferring 3D shapes without supervision by "inverting" the physics of image formation.

In conclusion, several technological and social challenges need to be addressed for the wide-scale adoption of AI systems. My research addresses several key ones: *generalization, robustness* and *interpretability* of computer vision systems. I am looking forward to continue innovating in these areas with the goal of building systems that can extract rich information from sensory inputs and act intelligently. I am also looking forward to collaborations with domain experts to develop novel applications and to advance science.

References

[1] Alessandro Achille, Michael Lam, Rahul Tewari, Avinash Ravichandran, Subhransu Maji, Charless Fowlkes, Stefano Soatto, and Pietro Perona. Task2Vec: Task Embedding for Meta-Learning. In *International Conference on Computer Vision (ICCV)*, 2019.

[2] Ejaz Ahmed, Gregory Shakhnarovich, and Subhransu Maji. Knowing a Good HOG Filter when You See It: Efficient Selection of Filters for Detection. In *European Conference on Computer Vision (ECCV)*, 2014.

[3] Lubomir Bourdev, Subhransu Maji, Thomas Brox, and Jitendra Malik. Detecting People using Mutually Consistent Poselet Activations. In *European Conference on Computer Vision (ECCV)*, 2010.

[4] Lubomir Bourdev, Subhransu Maji, and Jitendra Malik. Describing People: Poselet-Based Approach to Attribute Classification. In *International Conference on Computer Vision (ICCV)*, 2011.

[5] Thomas Brox, Lubomir Bourdev, Subhransu Maji, and Jitendra Malik. Object Segmentation by Alignment of Poselet Activations to Image Contours. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[6] Zezhou Cheng, Saadia Gabriel, Pankaj Bhambhani, Daniel Sheldon, Subhransu Maji, Andrew Laughlin, and David Winkler. Detecting and tracking communal bird roosts in weather radar data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 378–385, 2020.

[7] Zezhou Cheng, Matheus Gadelha, Subhransu Maji, and Daniel Sheldon. A Bayesian Perspective on the Deep Image Prior. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[8] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing Textures in the Wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[9] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, and Andrea Vedaldi. Deep Filter Banks for Texture Recognition, Description, and Segmentation. *International Journal of Computer Vision (IJCV)*, January 2016.

[10] Mircea Cimpoi, Subhransu Maji, and Andrea Vedaldi. Deep Filter Banks for Texture Recognition and Description. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[11] Matheus Gadelha, Subhransu Maji, and Rui Wang. 3D Shape Induction from 2D Views of Multiple Objects. In *International Conference on 3D Vision (3DV)*, 2017.

[12] Matheus Gadelha, Aartika Rai, Subhransu Maji, and Rui Wang. Inferring 3D Shapes from Image Collections using Adversarial Networks. *arXiv preprint*, abs/1906.04910, June 2019.

[13] Matheus Gadelha, Rui Wang, and Subhransu Maji. Multiresolution tree networks for 3d point cloud processing. In *The European Conference on Computer Vision (ECCV)*, September 2018.

[14] Matheus Gadhela, Subhransu Maji, and Rui Wang. 3D Shape Generation using Spatially Ordered Point Clouds. In *British Machine Vision Conference (BMVC)*, 2017.

[15] Tamir Hazan, Subhransu Maji, and Tommi Jaakkola. On Sampling from the Gibbs Distribution with Random Maximum A-Posteriori Perturbations. In *Neural Information Processing Systems (NIPS)*, 2013.

[16] Tamir Hazan, Subhransu Maji, Joseph Keshet, and Tommi Jaakkola. Learning Efficient Random Maximum A-Posteriori Predictors with Non-Decomposable Loss Functions. In *Neural Information Processing Systems (NIPS)*, 2013.

[17] Tamir Hazan, Francesco Orabona, Anand D. Sarwate, Subhransu Maji, and Tommi Jaakkola. High dimensional inference with random maximum a-posteriori perturbations. *IEEE Transactions on Information Theory*, 65, 2019.

[18] Kyle G Horton, Frank A La Sorte, Daniel Sheldon, Tsung-Yu Lin, Kevin Winner, Garrett Bernstein, Subhransu Maji, Wesley M Hochachka, and Andrew Farnsworth. Phenology of nocturnal avian migration has shifted at the continental scale. *Nature Climate Change*, 10(1):63–68, 2020.

[19] Evangelos Kalogerakis, Melinos Averkiou, Subhransu Maji, and Siddhartha Chaudhuri. 3D shape segmentation with projective convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[20] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran, and Stefano Soatto. Meta-learning with differentiable convex optimization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[21] Stephen Lee, Srinivasan Iyengar, Menghong Feng, Prashant Shenoy, and Subhransu Maji. DeepRoof: A Data-driven Approach For Solar Potential Estimation Using Rootop Imagery. In *SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2019.

[22] Tsung-Yu Lin and Subhransu Maji. Visualizing and Understanding Deep Texture Representations. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[23] Tsung-Yu Lin and Subhransu Maji. Improved Bilinear Pooling with CNNs. In *British Machine Vision Conference (BMVC)*, 2017.

[24] Tsung-Yu Lin, Subhransu Maji, and Piotr Koniusz. Second-order Democratic Pooling. In *European Conference on Computer Vision (ECCV)*, 2018.

[25] Tsung-Yu Lin, Aruni RoyChowdhury, and Subhransu Maji. Bilinear CNN Models for Fine-grained Visual Recognition. In *International Conference on Computer Vision (ICCV)*, 2015.

[26] Tsung-Yu Lin, Aruni RoyChowdhury, and Subhransu Maji. Bilinear Convolutional Neural Networks for Fine-grained Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2017.

[27] Tsung-Yu Lin, Mikayla Timm, Chenyun Wu, and Subhransu Maji. Visualizing and describing fine-grained categories as textures. *arXiv preprint arXiv:1907.05288*, 2019.

[28] Tsung-Yu Lin, Kevin Winner, Garrett Bernstein, Abhay Mittal, Adriaan M. Dokter, Kyle G. Horton, Cecilia Nilsson, Benjamin M. Van Doren, Andrew Farnsworth, Frank A. La Sorte, Subhransu Maji, and Daniel Sheldon. MistNet: Measuring historical bird migration in the US using archived weather radar data and convolutional neural networks. *Methods in Ecology and Evolution*, 2019.

[29] Zhaoliang Lun, Matheus Gadelha, Evangelos Kalogerakis, Subhransu Maji, and Rui Wang. 3d shape reconstruction from sketches via multi-view convolutional networks. In *International Conference on 3D Vision (3DV)*, 2017.

[30] Subhransu Maji. Discovering a Lexicon of Parts and Attributes. In *Second International Workshop on Parts and Attributes, ECCV*, 2012.

[31] Subhransu Maji. *A Taxonomy of Part and Attribute Discovery Techniques*, pages 247–268. Springer International Publishing, 2017.

[32] Subhransu Maji and Alexander Berg. Max-Margin Additive Classifiers for Detection. In *International Conference on Computer Vision (ICCV)*, 2009.

[33] Subhransu Maji, Alexander Berg, and Jitendra Malik. Classification using Intersection Kernel Support Vector Machines is Efficient. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[34] Subhransu Maji, Alexander Berg, and Jitendra Malik. Efficient Classification for Additive Kernel SVMs. In *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 35, No. 1, January 2013.

[35] Subhransu Maji, Lubomir Bourdev, and Jitendra Malik. Action Recognition from a Distributed Representation of Pose and Appearance. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[36] Subhransu Maji, Tamir Hazan, and Tommi Jaakkola. Efficient Boundary Annotation using Random Maximum A-Posteriori Perturbations. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2014.

[37] Subhransu Maji and Jitendra Malik. Object Detection using a Max-margin Hough Transform. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[38] Subhransu Maji and Gregory Shakhnarovich. Part Discovery from Partial Correspondence. In *Computer Vision and Pattern Recognition (CVPR)*, 2013.

[39] Subhransu Maji, Nisheeth Vishnoi, and Jitendra Malik. Biased Normalized Cuts. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[40] Gustavo Pérez, Matteo Messa, Daniela Calzetti, Subhransu Maji, Dooseok E Jung, Angela Adamo, and Mattia Sirressi. Starcnet: Machine learning for star cluster identification. *The Astrophysical Journal*, 907(2):100, 2021.

[41] Soha Rostaminia, Alexander Lamson, Subhransu Maji, Tauhidur Rahman, and Deepak Ganesan. W!nce: Unobtrusive sensing of upper facial action units with eog-based eyewear. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. (UBICOMP)*, 3(1):23:1–23:26, Mar. 2019.

[42] Manolis Savva, Fisher Yu, Hao Su, M Aono, B Chen, D Cohen-Or, W Deng, Hang Su, Song Bai, Xiang Bai, et al. SHREC'16 track: Large-scale 3D shape retrieval from ShapeNet Core55. In *Proceedings of the Eurographics Workshop on 3D Object Retrieval*, 2016.

[43] Gopal Sharma, Rishabh Goyal, Difan Liu, Evangelos Kalogerakis, and Subhransu Maji. CSGNet: Neural Shape Parser for Constructive Solid Geometry. In *Computer Vision and Pattern Recognition (CVPR)*, 2018.

[44] Hang Su, Varun Jampani, Deqing Sun, Subhransu Maji, Evangelos Kalogerakis, Ming-Hsuan Yang, and Jan Kautz. SPLATNet: Sparse lattice networks for point cloud processing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2530–2539, 2018.

[45] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view Convolutional Neural Networks for 3D Shape Recognition. In *International Conference on Computer Vision (ICCV)*, 2015.

[46] Jong-Chyi Su, Matheus Gadelha, Rui Wang, and Subhransu Maji. A deeper look at 3d shape classifiers. In *Second Workshop on 3D Reconstruction Meets Semantics, ECCV*, 2018.

[47] Jong-Chyi Su and Subhransu Maji. Adapting models to signal degradation using distillation. In *British Machine Vision Conference (BMVC)*, 2017.

[48] Jong-Chyi Su, Chenyun Wu, Huaizu Jiang, and Subhransu Maji. Reasoning about fine-grained attribute phrases using reference games. In *International Conference on Computer Vision (ICCV)*, 2017.

[49] Mikayla Timm, Subhransu Maji, and Todd Fuller. Large-scale ecological analyses of animals in the wild using computer vision. In *Fine-Grained Visual Categorization Workshop (FGVC5)*, 2018.

[50] Andrea Vedaldi, Siddharth Mahendran, Stavros Tsogkas, Subhransu Maji, Ross Girshick, Juho Kannala, Esa Rahtu, Iasonas Kokkinos, Matthew B Blaschko, David Weiss, Ben Taskar, Karen Simonyan, Naomi Saphra, and Sammy Mohamed. Understanding Objects in Detail with Fine-grained Attributes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[51] Catherine Wah, Grant Van Horn, Steven Branson, Subhransu Maji, Pietro Perona, and Serge Belongie. Similarity Comparisons for Interactive Fine-Grained Categorization. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[52] Catherine Wah, Subhransu Maji, and Serge Belongie. Learning Localized Perceptual Similarity Metrics for Interactive Categorization. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2015.

[53] Yang Zhou, Zhan Xu, Chris Landreth, Evangelos Kalogerakis, Subhransu Maji, and Karan Singh. VisemeNet: Audio-Driven Animator-Centric Speech Animation. *ACM Transactions on Graphics*, 37(4), 2018.