Artifice: Generalized Object Detection in Scientific Images

Benjamin Killeen and Gordon Kindlmann

***Abstract*—Together with large training sets, Deep Neural Networks (DNNs) have enabled a wide range of advancements in Computer Vision tasks, especially with regard to object detection in real-world images. Object detection for scientific images, on the other hand, presents challenges unlike more traditional vision tasks. First, experiments exhibit a scarcity of labelled data, a problem which we address through active learning and data augmentation. Second, scientific experiments involve inherent constraints, such as a natural law or governing principle, which the experimenter intends to observe. Although this quality seems advantageous for data augmentation, it threatens to undermine the scientific rigor of DNNs for laboratory tasks. In this progress report, we describe our ongoing effort toward training DNNs which are agnostic to inherent constraints yet perform reliably on scarce data.[1]**

I. Introduction

Imaging systems are a vital component of many scientific experiments, used as the primary measuring device of properties such as object location and orientation. In cases where no alternative measuring device exists, accurate labeling is of the utmost importance. However, such image analysis can prove labor-intensive. Traditional methods involve *ad hoc* solutions suited for one experimental setup, even though more general object detection tasks have been well-studied in Computer Vision. [1] uses a graph-cut approach for dot localization, noting various scientific applications, and [2] applies a DNN for semantic segmentation of neuronal structures in electron microscope stacks. In this report, we introduce *Artifice*, an ongoing effort to generalize across scientific tasks using DNNs.

Deep Neural Networks (alternatively, deep convolutional neural networks) have shown remarkable success on real-world tasks [3], as defined by datasets like ImageNet [4] and COCO [5]. These and other data underlie supervised learning approaches for increasingly complex systems, such as self-driving vehicles. Fig. 1a shows a high-level outline of supervised learning for computer vision. Scientific experiments, on the other hand, have no dataset except what they generate; each experiment constitutes its own unique vision task. Artifice addresses this scarcity of labeled data through active learning [6], [7], [8] and data augmentation [3], [2]. It also confronts a problem unique to scientific tasks, which we refer to as inherent constraints.

Benjamin Killeen is an undergraduate with the Department of Computer Science, University of Chicago, Chicago, IL 60637, USA (email: killeen@uchicago.edu)

Gordon Kindlmann is with the Department of Computer Science, University of Chicago, Chicago, IL 60637, USA (email: glk@uchicago.edu)

[1]Code available at github.com/bendkill/artifice

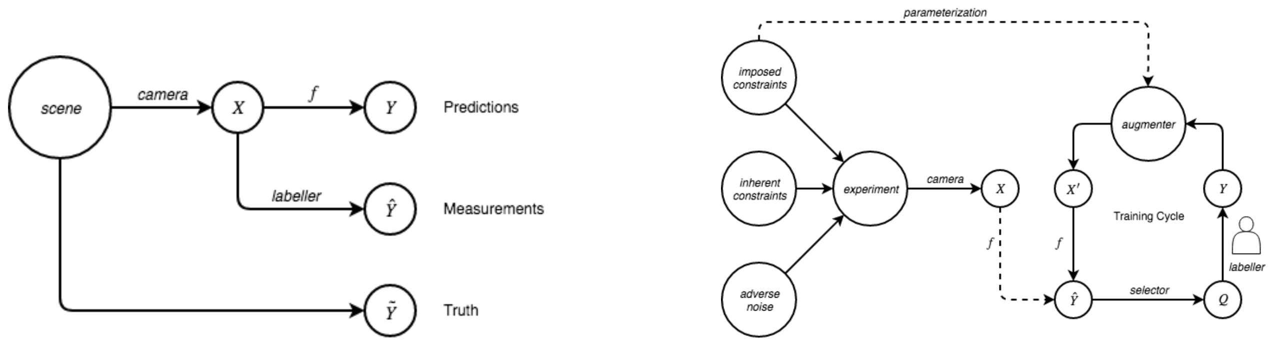
Unlike real-world tasks, scientific tasks aims to capture some natural law or governing principle in the data. Images of a sphere in free-fall, for example, capture the acceleration of gravity; a video of ants hunting for food obeys simple rules governing swarm behavior. These principles constitute *inherent constraints*: patterns in the dataset reflecting the subject of scientific inquiry itself. They differ from *imposed constraints*, which the experimenter controls. The experiment shown in fig. 2a, for example, constrains each moving dots to a circular region.

We distinguish between imposed and inherent constraints for the purpose of training. Because imposed constraints reduce the vision task, Artifice incorporates them into its data augmentation step, as in Fig. 1b. Crucially, imposed constraints are well-parameterized and guaranteed; the experimenter builds them into the vision task. Inherent constraints offer a similar advantage, since one could simulate their effects to further augment training data, but *doing so could violate the scientific method*. The experimenter cannot anticipate inherent constraints in order to build a better measuring stick, so to speak, because scientific experiments require unbiased measurement. Even the initial dataset, before any augmentation, contains inherent constraints that could influence training. As much as possible, Artifice aims to be agnostic to these constraints.

II. Related Work

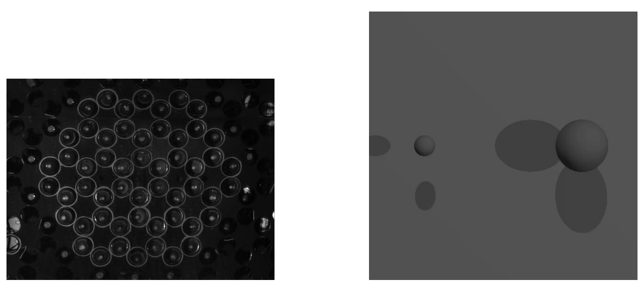
The problems of data scarcity and biases are well-considered in computer vision. [9] explores bias in popular datasets for computer vision by training DNNs on one dataset but testing them on another (employing a test-train split for fairness). Unsurprisingly, networks perform best on the dataset for which they were trained, even though datasets like ImageNet aim to capture the unbiased visual world. Creating unbiased datasets for real-world vision tasks remains an open problem, one which may demand a reckoning for the community's focus on improving dataset performance scores.

[2] describes a novel segmentation architecture using up-convolutions which, paired with their data augmentation scheme, performs well on biomedical images taken from electron stacks. Like Artifice, [2] confronts data scarcity but also focuses on applications in biomedical imaging, where semantic segmentation is often sufficient. Artifice aims to address scientific tasks more generally, using semantic segmentation as one component of a general detection system, capable of recovering location, orientation, or shape for multiple objects. Capturing these quantities with minimal human effort would accelerate a wide range of scientific experiments.



(a) (b)

Fig. 1. For a traditional vision task **(a)**, a camera converts real-world *scenes* into an unlabeled dataset X. These scenes have actual attributes $\tilde{Y}$ which a *labeller* records as the "ground truth" or "measurement" Y. A classifier or regression model f, trained on (X, Y), produces predictions $\hat{Y}$. Artifice's dependency graph **(b)**, specialized for scientific vision tasks, distinguishes between two types of constraints. *Imposed constraints*, such as physical boundaries, are known to the experimenter, but *inherent constraints*, such as a physical law, are the subject of inquiry. *Adverse noise*, including lighting changes or object deformation, also influences the experiment. In **(b)**, the augmented dataset X is continually updated by the Training Cycle. A *selector* chooses the query set $Q \subseteq X'$ based on existing predictions $\hat{Y}$. A *labeler* produces imperfect ground truth Y, which enables the *augmenter* to intelligently update the dataset X', incorporating imposed constraints. Here, dashed lines represent single instance connections, such as instantiating $X' \leftarrow X$, rather than continual dependency throughout the Training Cycle.



(a) Gyroscopic Model (b) Coupled Spheres

Fig. 2. Example images from scientific experiments: **(a)** gyroscopic model for topological metamaterials [10]. Each dot is constrained to the circle surrounding it. **(b)** still frame from a simulated experiment of two spheres coupled by an invisible spring. Full video here.

III. Simulated Experiments

In order to show Artifice's effectiveness, we develop several virtual experiments. These simulations offer several advantages over images from real experiments, such as in Fig. 2a. First, we have perfect knowledge of the experiment's "Truth" $\tilde{Y}$, as opposed to imperfect measurements, or "ground truth," Y. In well established datasets, $\tilde{Y}$ and Y are nearly identical, but Artifice must rely on one or a few human labelers for what should be unambiguous quantities. For testing, we calculate $\tilde{Y}$ from the known parameters of the simulation, and we emulate a human labeler by introducing small perturbations to $\tilde{Y}$, producing Y. Part of Artifice's goal is to train a DNN with predictions $\hat{Y}$ that more closely approximate $\tilde{Y}$ than the labels Y. Simulated experiments allow us to test this performance.

Fig. 2b shows one such experiment. In this case, two spheres with different masses rotate in free space, coupled by an invisible spring. The goal of the Coupled Spheres experiment is to recover physical properties of the spring using (x, y) positions of the two spheres. Imposed constraints include the z-coordinate of each sphere, which is set to the image-plane, as well as each sphere's apparent size. Inherent constraints include the physical properties of the spring, *e.g.* the spring constant and relaxed length. Adverse noise, which in this case includes any shadows, lighting effects, and possible occlusion, also presents a challenge for detection.

To demonstrate Artifice's resilience to inherent constraints, we intend to train a DNN on one experiment and evaluate its performance on experiments with different simulated springs. This simple example illustrates the general resilience that we wish to develop.

IV. Method

Fig. 1b shows Artifice's training scheme at a high-level. The details of this scheme are an area of active inquiry, with open questions pertaining to the *selector*, for active learning; the *augmenter*, which should incorporate the experiment's imposed constraints; and the model f, which employs one or more DNNs.

Because of the variety of scientific tasks, f must remain flexible with regard to its output. Toward this end, we envision a two-step procedure which (1) obtains an instance segmentation [2], [11] for objects of interest and (2) learns the target parameters (location, orientation, etc.) for each object. Whether these steps occur in an end-to-end fashion or are divided between several training steps is an open question. We do, however, consider (1) to be a vital step for the sake of maintaining generality. Even in scientific tasks where segmentation is unneeded, such as the Coupled Sphere experiment in IV, obtaining pixel-level masks of each object enables more advanced data augmentation methods.

Because we wish to minimize the man-hours required for training as much as possible, Artifice uses an active learning scheme to select a query set $Q \subseteq X'$ which will most inform training. The application of active learning to semantic segmentation has received relatively little attention, with the exception of [12], and the added complication of an imperfect labeler remains an open question in the field [6]. Artifice aims to address both issues with its *selector*.

Finally, Artifice's *augmenter* will incorporate both the semantic segmentations obtained by f and the imposed constraints specified by the experimenter to improve training as much as possible. We hope to test many augmentation methods while keeping in mind that the image space of a scientific task is usually much more constrained than that of a real-world task. [3], for instance, uses sub-image extraction, flipping, and PCA analysis of RGB channels to augment ImageNet. These methods are not necessarily applicable to scientific tasks, where imposed constraints might invalidate a flipped image, for instance.

V. Conclusion

We have outlined our ongoing efforts to develop Artifice, an object detection scheme intended for scientific tasks. Much work remains to be done with regard to Artifice's implementation, but we hope that by outlining our goals early, we have elucidated some key problems unique to scientific tasks.

Acknowledgment

Thanks to Michael Maire for input and guidance, as well as William Irvine for access to image data from his laboratory.

References

[1] Bernardis, E., and Yu, S. X., 2010. "Finding dots: Segmentation as popping out regions from boundaries". In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 199–206.

[2] Ronneberger, O., Fischer, P., and Brox, T., 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation". *arXiv:1505.04597 [cs]*, May. arXiv: 1505.04597.

[3] Krizhevsky, A., Sutskever, I., and Hinton, G. E., 2012. "ImageNet Classification with Deep Convolutional Neural Networks". In *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds. Curran Associates, Inc., pp. 1097–1105.

[4] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. "ImageNet: A Large-Scale Hierarchical Image Database". p. 8.

[5] Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollr, P., 2014. "Microsoft COCO: Common Objects in Context". *arXiv:1405.0312 [cs]*, May. arXiv: 1405.0312.

[6] Settles, B., 2012. "Active Learning". *Synthesis Lectures on Artificial Intelligence and Machine Learning,* **6**(1), June, pp. 1–114.

[7] Kao, C.-C., Lee, T.-Y., Sen, P., and Liu, M.-Y., 2018. "Localization-Aware Active Learning for Object Detection". *arXiv:1801.05124 [cs]*, Jan. arXiv: 1801.05124.

[8] Jain, S. D., and Grauman, K., 2016. "Active Image Segmentation Propagation". In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2864–2873.

[9] Torralba, A., and Efros, A. A., 2011. "Unbiased look at dataset bias". In CVPR 2011, pp. 1521–1528.

[10] Nash, L. M., Kleckner, D., Read, A., Vitelli, V., Turner, A. M., and Irvine, W. T. M., 2015. "Topological mechanics of gyroscopic metamaterials". *Proceedings of the National Academy of Sciences,* **112**(47), pp. 14495–14500.

[11] Bai, M., and Urtasun, R., 2016. "Deep Watershed Transform for Instance Segmentation". *arXiv:1611.08303 [cs]*, Nov. arXiv: 1611.08303.

[12] Vezhnevets, A., Buhmann, J. M., and Ferrari, V., 2012. "Active learning for semantic segmentation with expected change". In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3162–3169.