Task Discovery: A Learnability Exploration

Ajay Sohmshetty (collaboration with Amir Zamir)
CS230: Deep Learning

Background

- We've seen from Taskonomy [1] that establishing and leveraging relationships between tasks can build better, more generalizable models. These tasks have been hand-chosen (bias), and in reality there are a much larger number of tasks, and it's impractical to annotate all or a large number of them.
- This project aims to generate these tasks by defining a task and investigate a series of reduction mechanisms. The ultimate goal would be to apply these as well as other reduction mechanisms to filter down to a set of generated tasks that can be learned from other models, which can then be used for transfer learning purposes.

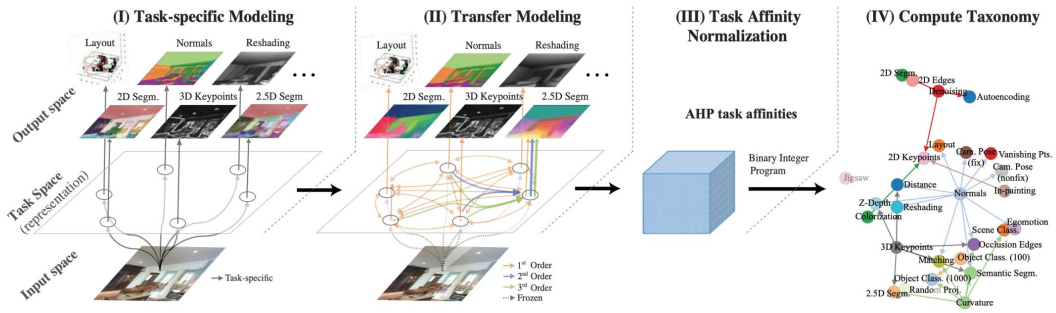

Figure 1: Overview of Taskonomy, which studied the clear structure between tasks and leverages them to reduce demand for labeled data.

Dataset and Architecture Used

- We pick a subset of CIFAR100 classes, and re-formulate the task into a binary classification problem.
- Use a shallow model to avoid excessive overfitting, since datasets themselves are quite small.

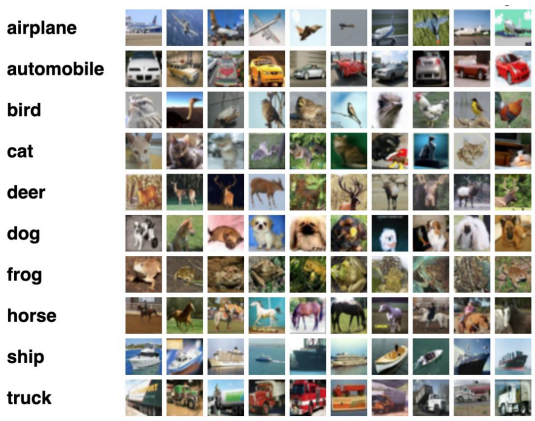

Figure 2: Sample CIFAR classes

#	Layer	Details
0	input	32x32x3
1	conv1	5x5 filters, 6 of them
2	pool	2x2 filter, stride 2
3	conv2	5x5 filters, 16 of them
4	fc1	120 hidden units
5	fc2	84 hidden units
6	fc3	2 output units

Table 1: SimpleNet Architecture

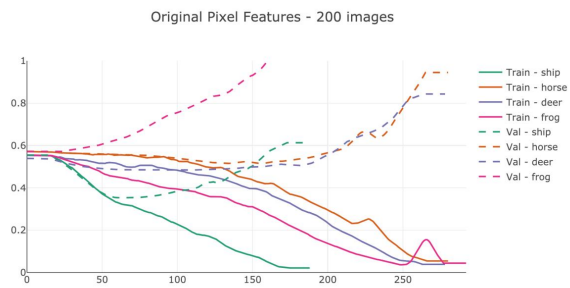

Figure 3: Sample run of SimpleNet on 4 randomly chosen classes.

Random Features

- Can we reduce dimensionality of our input space? If yes, we can work with larger images, scale down without incorporating any priors into model.
- Use randomly initialized ResNets as feature extractors - better than using pretrained networks or other feature extractors for our purposes.
- 3x8x8 feature shapes show promise (a 4x dimensionality savings)!

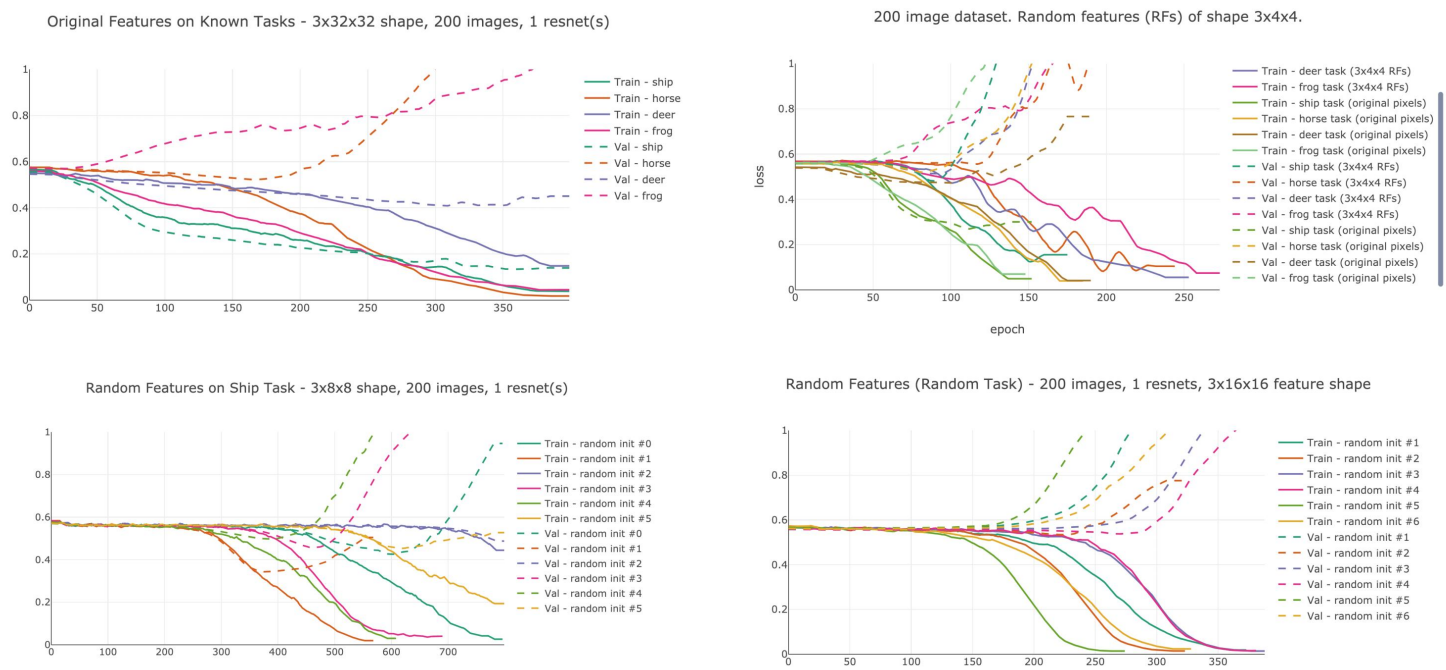

Figure 4: Random features exploration results.

Learnability Studies

- Ablation study
- Dataset size and class distribution study
- Fixed number of positive datapoints study
- Settle on 400 dataset size, 1:9 ratio.

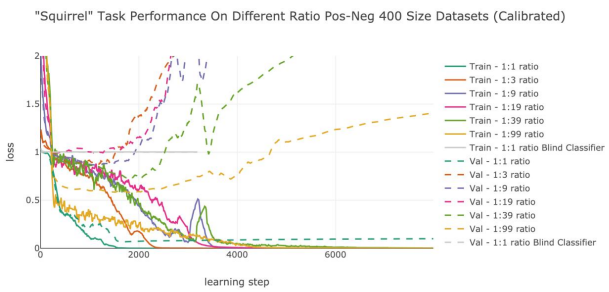

Figure 5: Class distribution analysis.

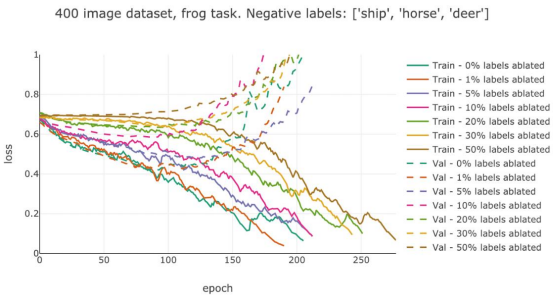

Figure 6: Ablation analysis for known task.

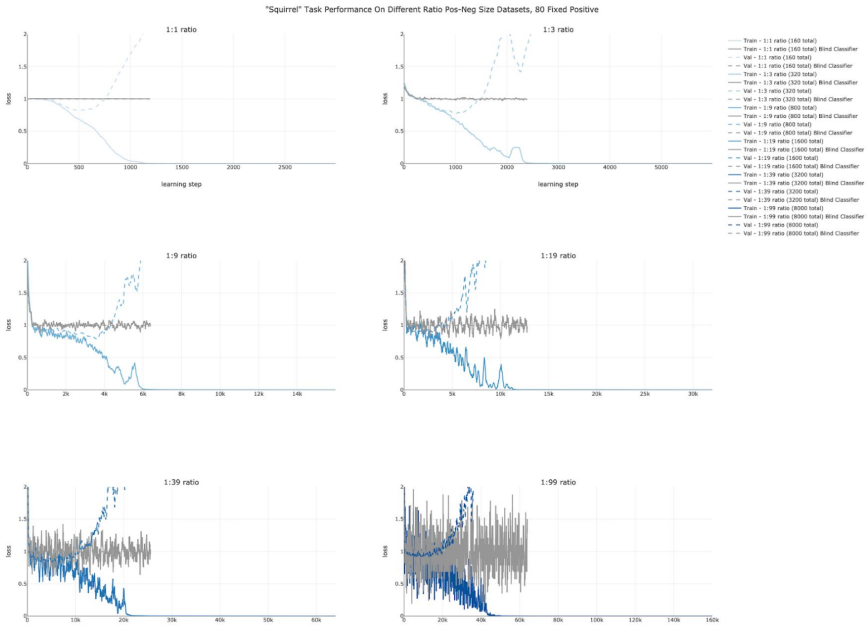

Figure 7: Fixed positive datapoints analysis.

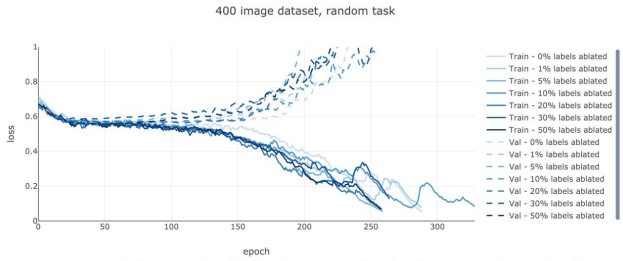

Figure 8: Ablation analysis for random task.

Task Discovery Search

- After clustering (both mean-shift and k-means) proved to be unfruitful in reducing task space, I instead formulated the random search into a gradient-free optimization problem, optimizing over “taskness score”.
- Initial results on smaller dataset (50 label set, 1:4 class distribution, 10 positive datapoints) look promising!

Figure 9: Run 1, taskness score 1-0.21/0.45 = 0.53

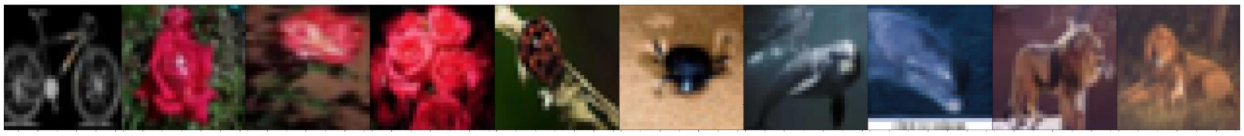

Figure 10: Run 2, taskness score 1-0.29/0.45 = 0.35

Figure 11: Run 3, taskness score 1-0.33/0.45 = 0.26

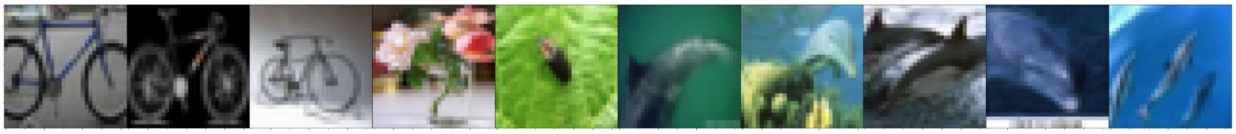

Figure 12: Run 4, taskness score 1-0.17/0.45 = 0.62

Conclusion

- We see encouraging results, as discovered task look meaningful and return real task-like taskness scores. The next logical step would be to perform a large scale task discovery search. This involves enforcing orthogonality among found tasks. I.e. we want to find sufficiently different tasks in each subsequent search. We hope to discover the original CIFAR100 tasks in this dataset, alongside other real, interesting tasks. If successful, we can leverage these tasks for transfer learning purposes. An automated method of discovering tasks such as the method presented in this paper may unlock the true potential of transfer learning. We hope to submit to NeurIPS this upcoming May.

References

1. http://taskonomy.stanford.edu/taskonomy_CVPR2018.pdf
2. http://www.cs.cmu.edu/~aarti/SMLRG/miguel_slides.pdf
3. https://code.fb.com/ai-research/nevergrad/
4. https://arxiv.org/pdf/1703.03864.pdf
5. https://arxiv.org/abs/1603.08155
6. https://people.eecs.berkeley.edu/~brecht/papers/07.rah.rec.nips.pdf
7. https://www.jstor.org/stable/2241175?seq=1#page_scan_tab_contents
8. https://ieeexplore.ieee.org/document/7258343/
9. https://distill.pub/2018/differentiable-parameterizations/
10. https://arxiv.org/abs/1512.03385