FINE-GRAINED OBJECT RECOGNITION USING WYZE CAM

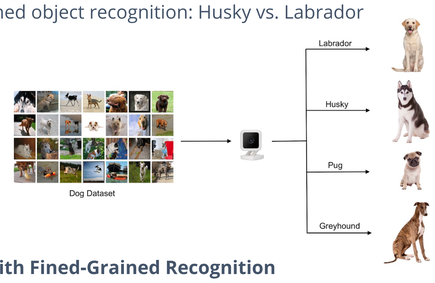
HSIANG-WEI HUANG, CONOR KNOX, TRINH NGUYEN, PRATHIBHA RAMACHANDRAN, RAVI SANGANI

44

Introduction

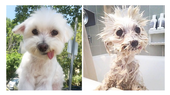
Object recognition is a computer vision problem for identifying objects in images or videos. There are two types of object recognition:

- Generic object recognition: dog vs. cat
- Fine-grained object recognition: Husky vs. Labrador



Challenges With Fined-Grained Recognition

Large intra-class variability



Large intra-class variability

Labeled data scarcity

Classifying mixed breeds

Modeling Solutions

Dataset

Stanford Dog Dataset: public dataset for training purposes
- Images of 120 different breeds for a total of 20,580 images

Our Datasets: custom datasets for testing purposes
- 3 datasets, each differs in image complexity
- Images of 25 different breeds for a total of 3,063 images

simple → complex

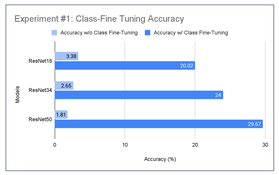
Google	Youtube	Real World
676 images	500 images	1887 images

Results

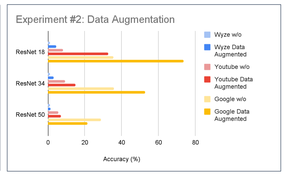
Phase 1: Architecture Evaluation Results

Model	Accuracy on Stanford Dog Dataset (12,000 images, 120 breeds)	Accuracy on a subset of our Google Dataset (88 images, 10 breeds)	Accuracy on a subset of our YouTube Dataset (88 images, 10 breeds)	Accuracy on a subset of our Real-World Dataset (88 images, 10 breeds)
VGG16	66.2%	55.7%	20.5%	9.6%
VGG19	66.8%	59.1%	22.7%	9.6%
ResNet50	80.4%	81.8%	58.0%	11.4%
DenseNet121	75%	81.8%	60.2%	9.1%
Mobilenetv2	65.9%	72.7%	35.2%	6.8%
NASNetMobile	74.4%	75.0%	28.4%	9.6%
EfficientNet B0	77.1%	78.4%	56.8%	9.1%
TransFG	92.3%	81.8%	50.0%	22.7%

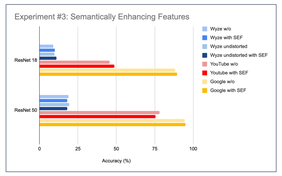
Phase 2: Flat Classification Experiment Results



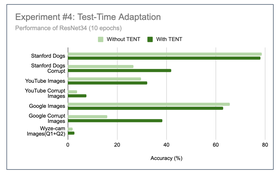
- **Fine-tuning on just 25 breeds** improved performance

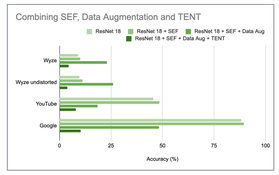


- **Data augmentation** significantly improved performance across majority of the models



- **SEF** provided modest increase in performance while being versatile and computationally cheap

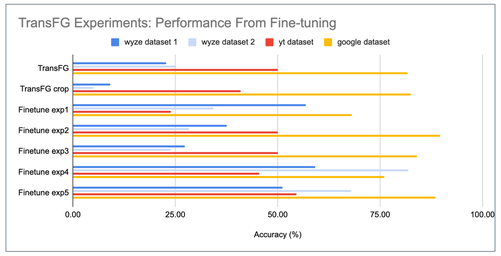




- **TENT** increased accuracy on corrupted images, but made negligible impact on Real-World data

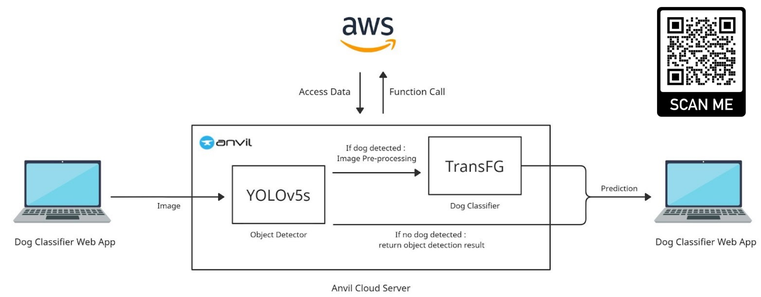
Phase 2: Fine-Grained Classification Experiment Results

- Different settings for the TransFG model fine-tuning experiments are conducted



Model	Pre-trained Dataset	Dataset	Epochs	Learning Rate
TransFG	ImageNet 21k	Stanford Dogs	10000	3.00E-02
TransFG crop	ImageNet 21k	Stanford Dogs Crop	10000	3.00E-02
Finetune exp1	ImageNet 21k, Stanford Dogs	Wyze Dataset V1	1000	3.00E-02
Finetune exp2	ImageNet 21k, Stanford Dogs	Wyze Dataset V2	100	3.00E-02
Finetune exp3	ImageNet 21k, Stanford Dogs	Wyze Dataset V2	100	3.00E-03
Finetune exp4	ImageNet 21k, Stanford Dogs	Wyze Dataset V2	1000	3.00E-02
Finetune exp5	ImageNet 21k, Stanford Dogs	Wyze Dataset V2	1000	3.00E-03

Technical Solution Architecture



Fine-Grained Object Tracker

- All data collected with Wyze Cam are video based
- Image-based classifier performed poorly when evaluating complex images collected from the real world
- Incorporated a simple tracking algorithm to conduct frame-by-frame association for every detection in the videos
- Majority vote classification is conducted after tracking to increase performance and robustness of classifier

Conclusion and Future Work

Researched fine–grained image recognition methods, created custom datasets, and successfully implemented a full-stack web application that can detect and classify 120 kinds of dog species.

Future Work

- Perform more experiments in fine-tuning to prevent overfitting and forgetting issues
- Enable pipeline to conduct fine-grained object recognition in more categories
- Explore real-time tracking for fine-grained object tracker
- Incorporate fine-grained recognition into Wyze Cam services

References

[1] He, Ju, et al. "TransFG: A Transformer Architecture for Fine-Grained Recognition." ArXiv.org, 1 Dec. 2021.

[2] He, Kaiming, et al. "Deep Residual Learning for Image Recognition." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016, pp. 770–778, 10.1109/cvpr.2016.90.

[3] Huang, Gao, et al. "Densely Connected Convolutional Networks." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[4] Hu, Tao, et al. "See Better before Looking Closer: Weakly Supervised Data Augmentation Network for Fine-Grained Visual Classification." ArXiv:1901.09891 [Cs], 23 Mar. 2019, arxiv.org/abs/1901.09891.

[5] Luo, Wei, et al. "Learning Semantically Enhanced Feature for Fine-Grained Image Classification." ArXiv:2006.13457 [Cs], 26 Aug. 2020, arxiv.org/abs/2006.13457.

[6] Simonyan, Karen, and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." ArXiv.org, 2014, arxiv.org/abs/1409.1556.

[7] Sun, Guolei, et al. "Fine-Grained Recognition: Accounting for Subtle Differences between Similar Classes." Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 07, 3 Apr. 2020, pp. 12047–12054, arxiv.org/pdf/1912.06842v1.pdf, 10.1609/aaai.v34i07.6882. Accessed 27 Apr. 2022.

[8] Tan, Mingxing, and Quoc V Le. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." ArXiv.org, 2019, arxiv.org/abs/1905.11946.

[9] Wei, Xiu-Shen, et al. "Deep Learning for Fine-Grained Image Analysis: A Survey." ArXiv:1907.03069 [Cs], 5 July 2019, arxiv.org/abs/1907.03069.

ELECTRICAL & COMPUTER ENGINEERING
UNIVERSITY of WASHINGTON

UW ADVISOR: ARKA MAJUMDAR
INDUSTRY ADVISORS: ZHONGWEI CHENG, PRADEEP SINGH
SPONSORS: WYZE LABS, INC