# Automated Classification of Model Errors on ImageNet

*Momchil Peychev\*, Mark Niklas Müller\*, Marc Fischer, Martin Vechev*
\*Equal contribution
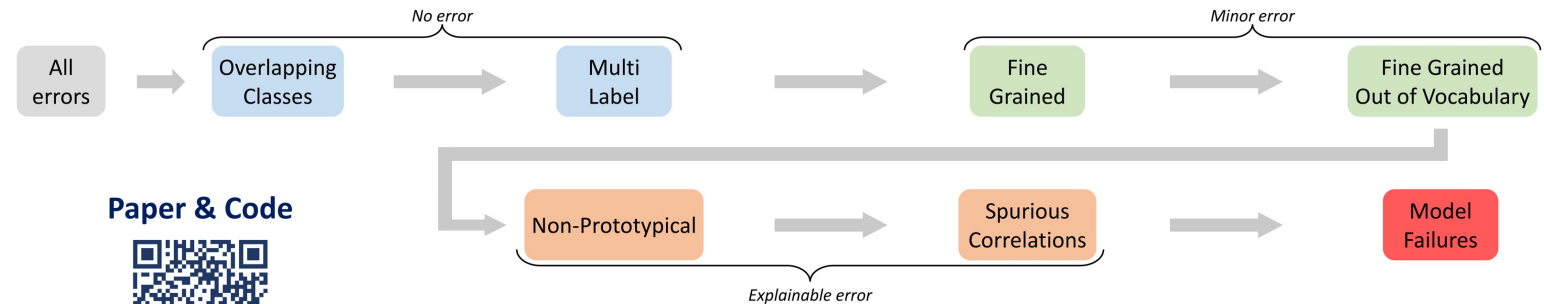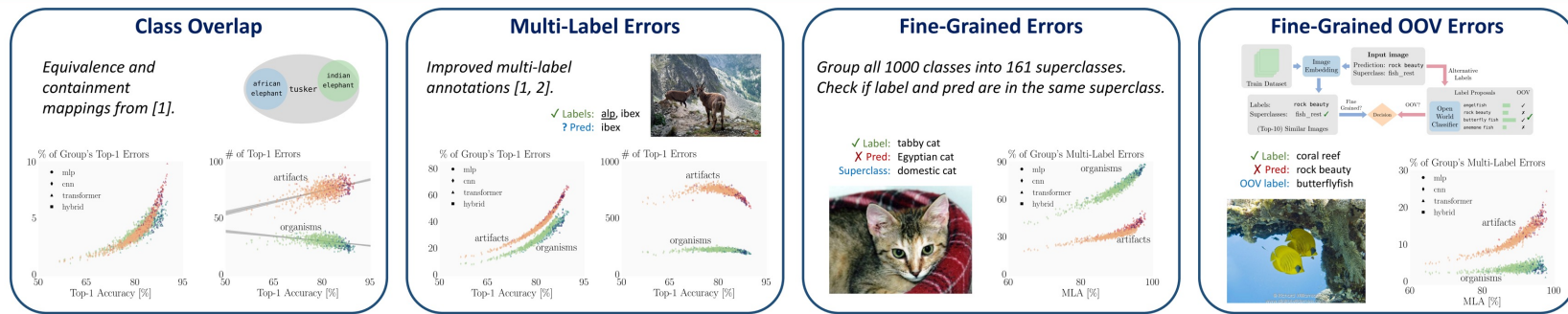
We analyze the mistakes of **900+** ImageNet classifiers.

Top-1 accuracy still matters.

Organisms get confused and artifacts are correlated or out-of-vocabulary.

- Top-1 and MLA are pessimistic measures of model progress.

- Organisms (41% of the classes) and artifacts (52.2%) exhibit very different error patterns.

## Paper & Code

SRILAB

NEURAL INFORMATION PROCESSING SYSTEMS

1. Vasudevan et al., "When does dough become a bagel? Analyzing the remaining mistakes on ImageNet", NeurIPS 2022
2. Shankar et al., "Evaluating Machine Accuracy on ImageNet", ICML 2020
3. Beyer et al., "Are we done with ImageNet?", arXiv 2020

### Class Overlap
*Equivalence and containment mappings from [1].*

african elephant · tusker · indian elephant

% of Group's Top-1 Errors
- mlp
- cnn
- transformer
- hybrid

# of Top-1 Errors
artifacts
organisms

Top-1 Accuracy [%]

### Multi-Label Errors
*Improved multi-label annotations [1, 2].*

✓ Labels: alp, ibex
? Pred: ibex

% of Group's Top-1 Errors
- mlp
- cnn
- transformer
- hybrid

# of Top-1 Errors
artifacts
organisms

Top-1 Accuracy [%]

### Fine-Grained Errors
*Group all 1000 classes into 161 superclasses. Check if label and pred are in the same superclass.*

✓ Label: tabby cat
✗ Pred: Egyptian cat
Superclass: domestic cat

% of Group's Multi-Label Errors
organisms
- mlp
- cnn
- transformer
- hybrid
artifacts

MLA [%]

### Fine-Grained OOV Errors

Train Dataset — Image Embedding — Input image
Prediction: rock beauty
Superclass: fish_rest
Alternative Labels

Labels: rock beauty
Superclasses: fish_rest
(Top-10) Similar Images

OOV — Open World Classifier — OOV

Label Proposals

✓ Label: coral reef
✗ Pred: rock beauty
OOV label: butterflyfish

% of Group's Multi-Label Errors
- mlp
- cnn
- transformer
- hybrid
artifacts
organisms

MLA [%]

### Pipeline

All errors → Overlapping Classes → Multi Label → Fine Grained → Fine Grained Out of Vocabulary

*No error*  *Minor error*

→ Non-Prototypical → Spurious Correlations → Model Failures

*Explainable error*

### Non-Prototypical Instances
*Categorization of non-prototypical images by [1].*

✓ Label: whistle
✗ Pred: perfume

% of Group's Multi-Label Errors
- mlp
- cnn
- transformer
- hybrid

# of Multi-Label Errors
artifacts
organisms

MLA [%]

### Spurious Correlations
*Extract pairs of commonly co-occuring classes from ImageNet-ReaL [3]. Check if a label and pred often co-occur together.*

✓ Labels: ski mask, alp
✗ Pred: ski

% of Group's Multi-Label Errors
- mlp
- cnn
- transformer
- hybrid
artifacts
organisms

MLA [%]

### Model Failures
*Left with the severe, unexplainable errors.*

✓ Labels: basket, hamper
✗ Pred: pillow

% of Group's Multi-Label Errors
artifacts
- mlp
- cnn
- transformer
- hybrid
organisms

MLA [%]

% of Group's Top-1 Errors
artifacts
- mlp
- cnn
- transformer
- hybrid
organisms

Top-1 Accuracy [%]