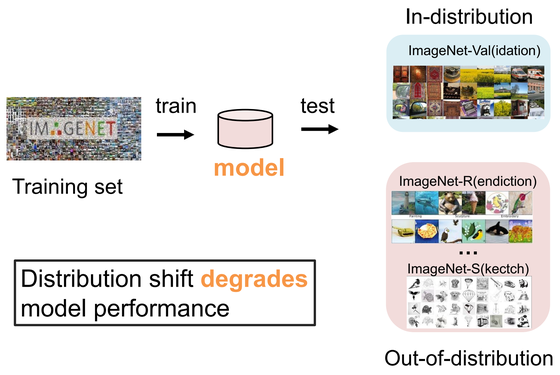
On the Strong Correlation Between Model Invariance and Generalization

Weijian Deng Stephen Gould Liang Zheng
Australian National University

Model Generalization

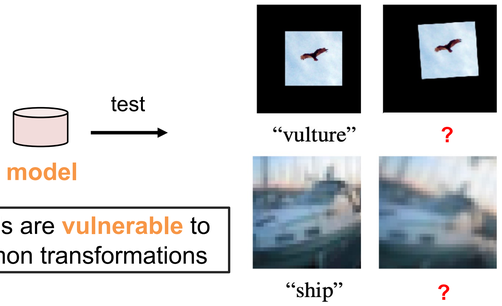
Generalization captures a model's ability to **classify unseen** data



Distribution shift **degrades** model performance

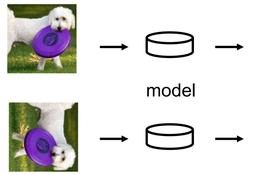
Model Invariance

Invariance measures how **consistent** the model predictions are on transformed test data



Models are **vulnerable** to Common transformations

Effective Invariance



A model should have a **high invariance** if and only if it is **highly confident** in predicting the **same class**

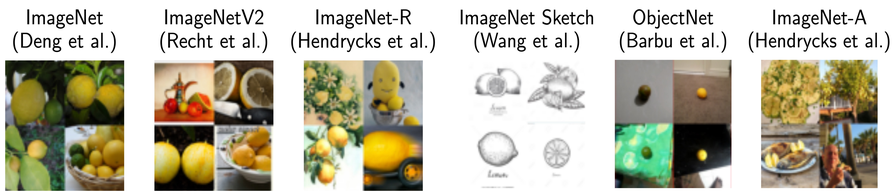
$$\mathrm{EI}(\boldsymbol{x}, \mathcal{T}(\boldsymbol{x}), \boldsymbol{f}) = \begin{cases} \sqrt{\hat{p}_t \cdot \hat{p}} & \text{if } \hat{y}_t = \hat{y}\,; \\ 0 & \text{otherwise}\,. \end{cases}$$

Geometric mean to consider the confidence scores

Correlation Study

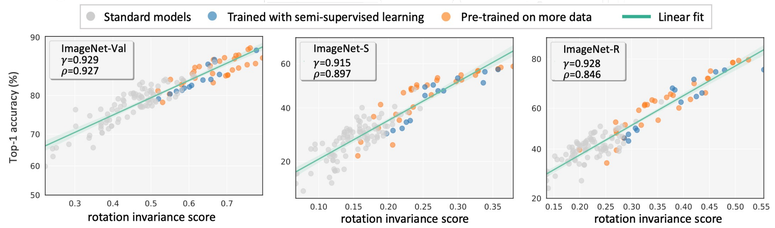
- 8 Test datasets

Style Shift, Dataset reproduction shift, Sketch shift, Blur shift, Natural adversarial shift, Bias-controlled shift …

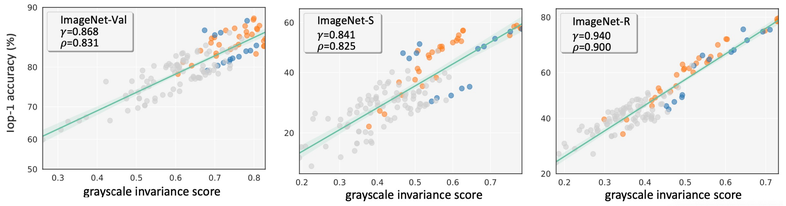


- Over 150 models
  1) Standard neural networks (*e.g.*, VGG and EfficientNet)
  2) Semi-supervised Learning (*e.g.*, SWSL-ResNet and EfficientNet-L2-NS)
  3) Pretraining on more data (*e.g.*, BiT and ViT-L)

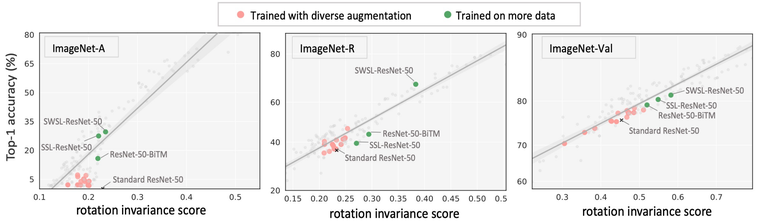
- **Strong Linear Correlation (Rotation Invariance)**



- **Strong Linear Correlation (Grayscale Invariance)**



- **Train on More Data vs. Data Augmentation**



- **Dataset Centric: A Model's Accuracy and Invariance on Many Test Sets**