Australian National University

NEURAL INFORMATION PROCESSING SYSTEMS

On the Strong Correlation Between Model Invariance and Generalization

Weijian Deng Stephen Gould Liang Zheng

Australian National University

Two Properties of a Machine Learning Model

Generalization

Classification ability on unseen data

Invariance

Robustness to input transformation

Generalization

Generalization captures a model's ability to classify unseen data

Generalization

Generalization captures a model's ability to classify unseen data

In-distribution Generalization

ImageNet-Val(idation)

ImageNet Training Set

ViT

Accuracy?

Generalization

Generalization captures a model's ability to classify unseen data

ImageNet Training Set

ViT

Accuracy?

In-distribution Generalization

ImageNet-Val(idation)

Out-of-distribution Generalization

ImageNet-R(endiction)

…

ImageNet-S(kectch)

Generalization

Generalization captures a model’s ability to classify unseen data

In-distribution Generalization

ImageNet-Val(idation)

Model often performs poorly on datasets that have a different distribution from that of the training data

ImageNet Training Set

ViT

...

ImageNet-S(kectch)

Invariance

Invariance measures how consistent model predictions are on transformed test data

Invariance

Invariance measures how consistent model predictions are on transformed test data

Invariance

Invariance measures how consistent model predictions are on transformed test data

State-of-the-art models turn out to be vulnerable to translations and rotations

Engstrom, Logan, et al. "Exploring the landscape of spatial robustness." In ICML, 2019.

Relationship Between Invariance and Generalization

Generalization

Classification ability on unseen data

Invariance

Robustness to input transformation

Relationship

Insights From Existing Works

“adding rotation invariance to a model improves its in-distribution (ID) classification accuracy”

*Zhou, Yanzhao, et al. "**Oriented response networks**." CVPR, 2017*
*Delchevalerie, Valentin, et al. "**Achieving Rotational Invariance with Bessel-Convolutional Neural Networks**.” NeurIPS, 2021*
*Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. "**Spatial transformer networks**." NIPS 2015.*

Relationship

Insights From Existing Works

"adding rotation invariance to a model improves its in-distribution (ID) classification accuracy"

*Zhou, Yanzhao, et al. "**Oriented response networks**." CVPR, 2017*
*Delchevalerie, Valentin, et al. "**Achieving Rotational Invariance with Bessel-Convolutional Neural Networks**." NeurIPS, 2021*
*Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. "**Spatial transformer networks**." NIPS 2015.*

"a shift-invariant model is robust to perturbation"

Zhang, Richard. "**Making convolutional networks shift-invariant again.**" ICML, 2019

Relationship

Insights From Existing Works

"adding rotation invariance to a model improves its in-distribution (ID) classification accuracy"

*Zhou, Yanzhao, et al. "**Oriented response networks**." CVPR, 2017*
*Delchevalerie, Valentin, et al. "**Achieving Rotational Invariance with Bessel-Convolutional Neural Networks**." NeurIPS, 2021*
*Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. "**Spatial transformer networks**." NIPS 2015.*

"a shift-invariant model is robust to perturbation"

Zhang, Richard. "**Making convolutional networks shift-invariant again.**" ICML, 2019

"theoretical investigations suggest that learning invariant features benefits model generalization"

Zhu, Sicheng, Bang An, and Furong Huang. "**Understanding the Generalization Benefit of Model Invariance from a Data Perspective**." *NeurIPS, 2021*

Relationship

Relationship Between Invariance and Generalization

Generalization

Invariance

Strong connection?

- Quantitative study

How to Measure?

Generalization → classification accuracy

Invariance → ?

Effective Invariance

1) Predicted class

2) Prediction confidence

original

model

$\hat{y}$

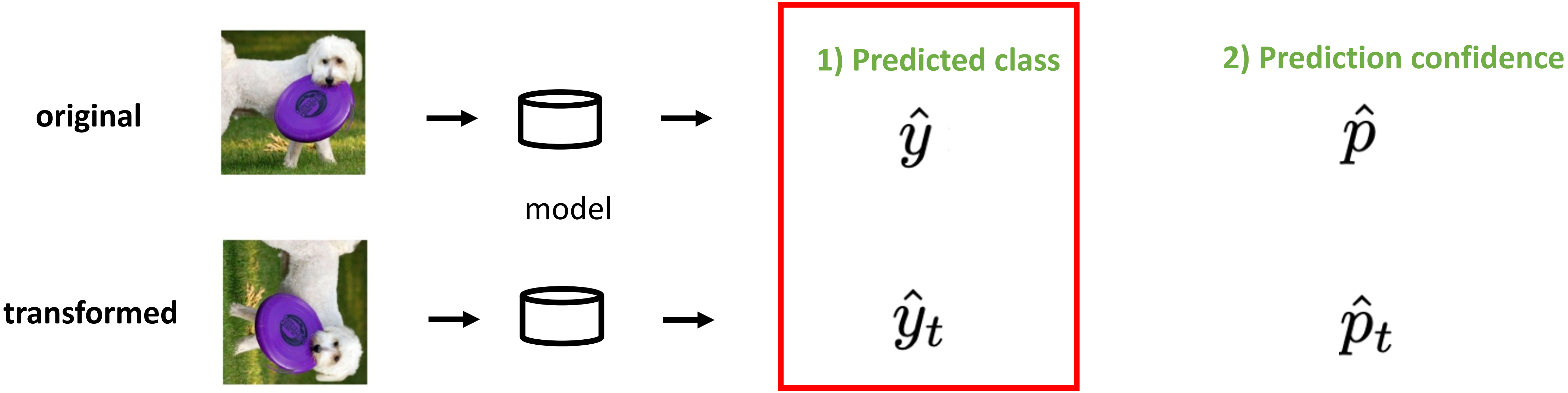
$\hat{p}$

transformed

$\hat{y}_t$

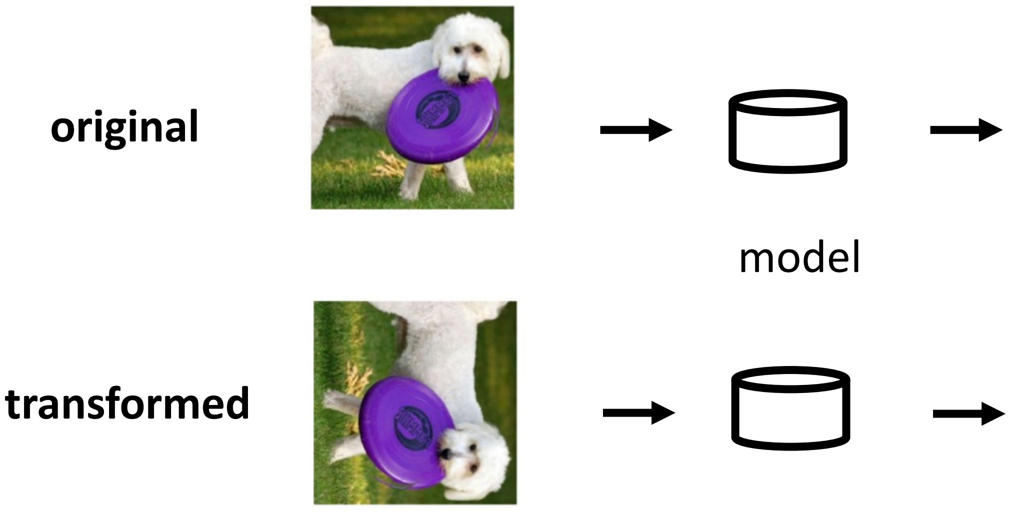
$\hat{p}_t$

Effective Invariance



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

Effective Invariance



1) Predicted class

$\hat{y}$

$\hat{y}_t$

2) Prediction confidence

$\hat{p}$

$\hat{p}_t$

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

Correlation Study: Models

ImageNet Setup

150 ImageNet models that are trained or finetuned on ImageNet training set

1) Standard neural networks: 100 models ***only*** trained on ImageNet training set

2) Semi-supervised learning: 15 models trained using a **large** collection of ***unlabelled*** images (*e.g.*, Instagram 900M)

2) Pretraining on more data: 35 models that are ***pre-trained*** on significantly **larger datasets** (*e.g.,* ImageNet-21K)

Correlation Study: Test Sets

ImageNet Setup

6 types of data distribution

In-Distribution: ImageNet-Val(idation)

Dataset reproduction shift: ImageNet-V2-A/B/C

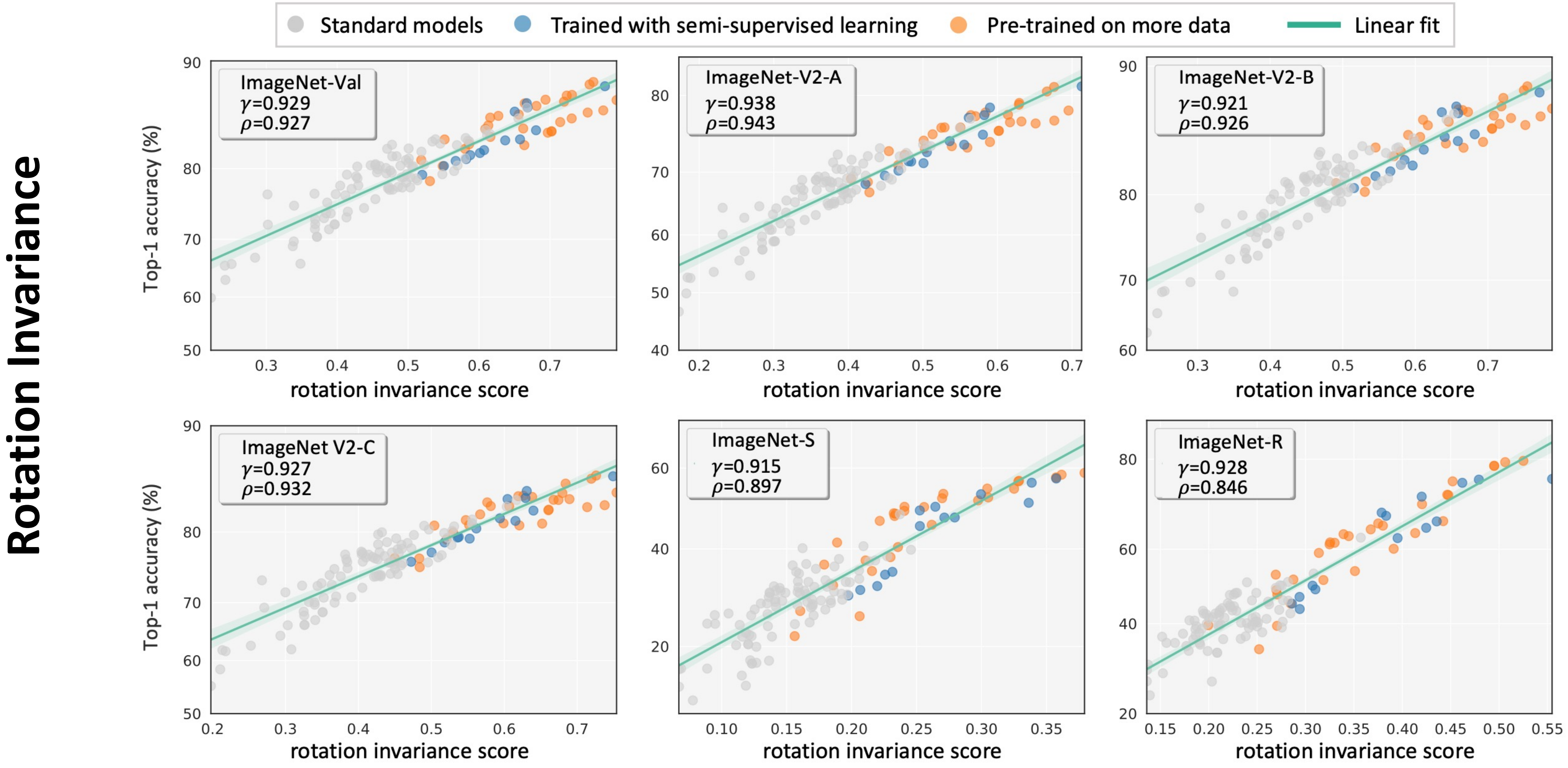
Natural adversarial shift: ImageNet-Adv(ersarial)

Sketch shift: ImageNet-S(ketch)

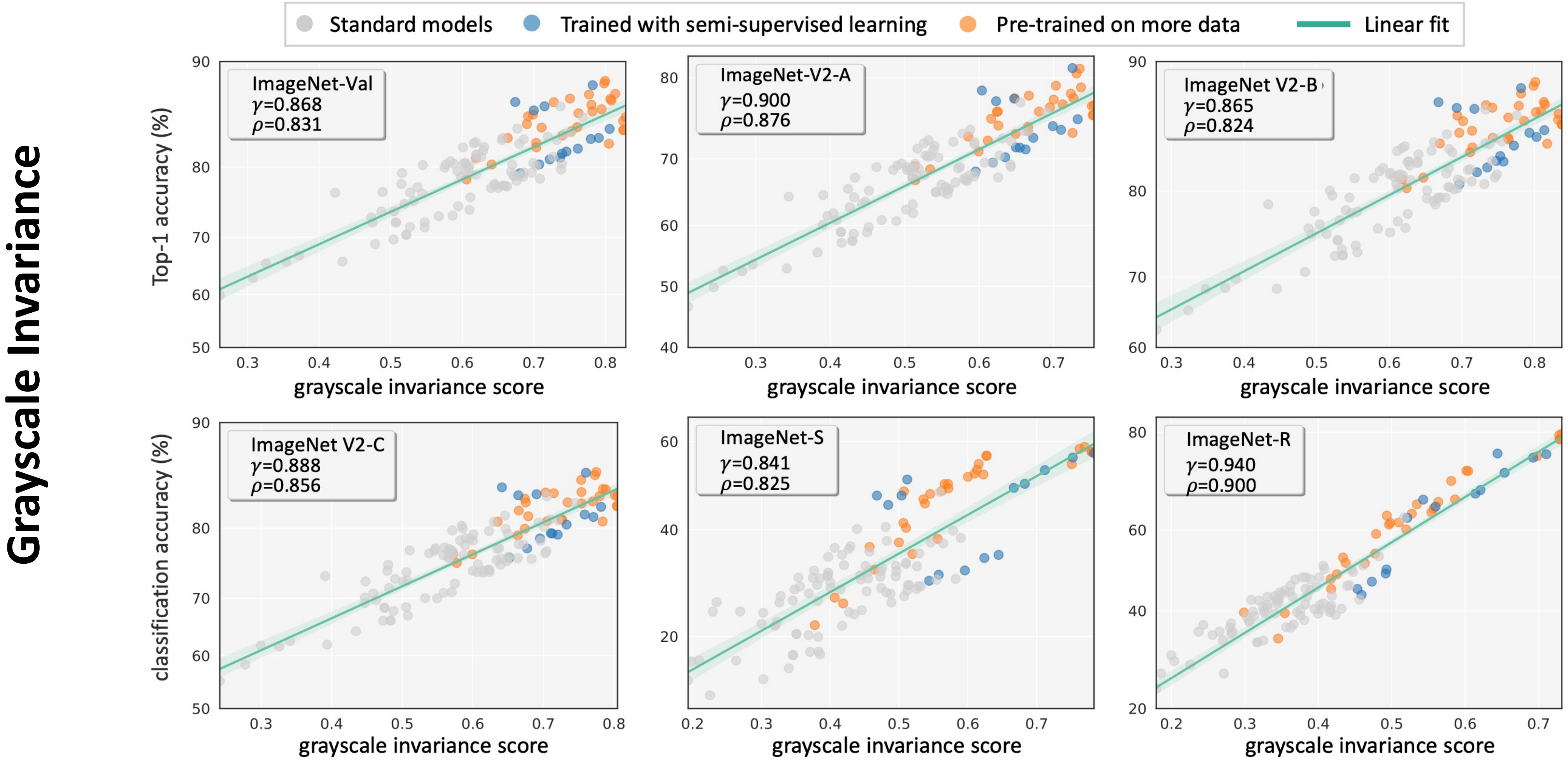
Blur shift: ImageNet-Blur that is synthesized by blurring ImageNet-Val

Style shift: ImageNet-R(endition)

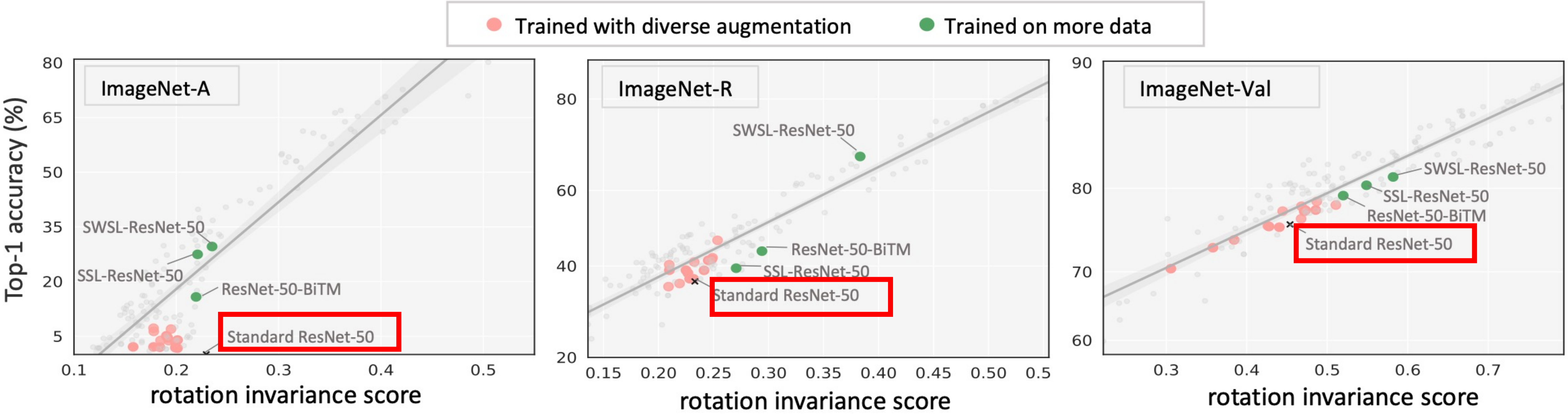
Correlation Study: Strong Correlation



Correlation Study: Strong Correlation



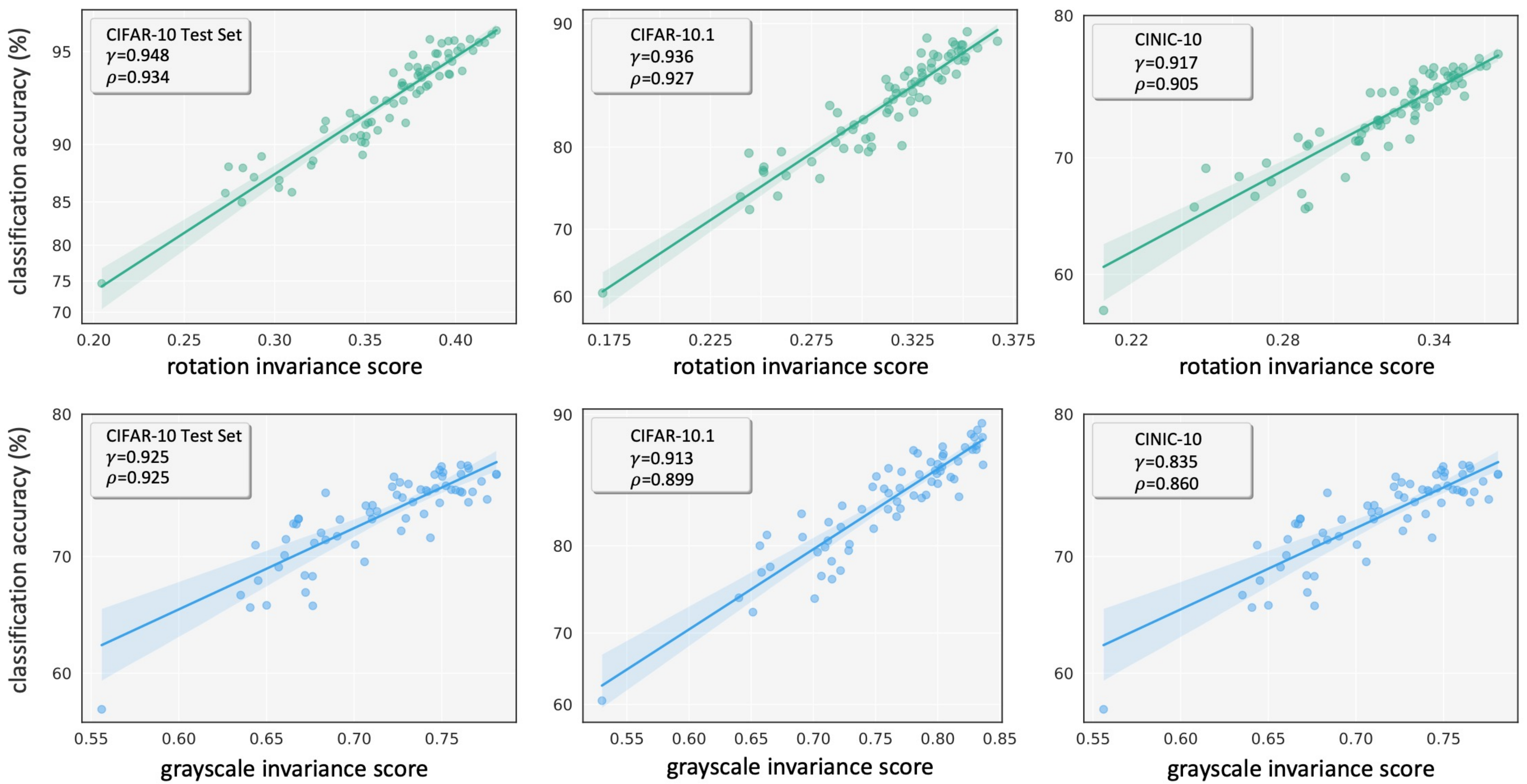
Data Augmentation vs. More Training Data



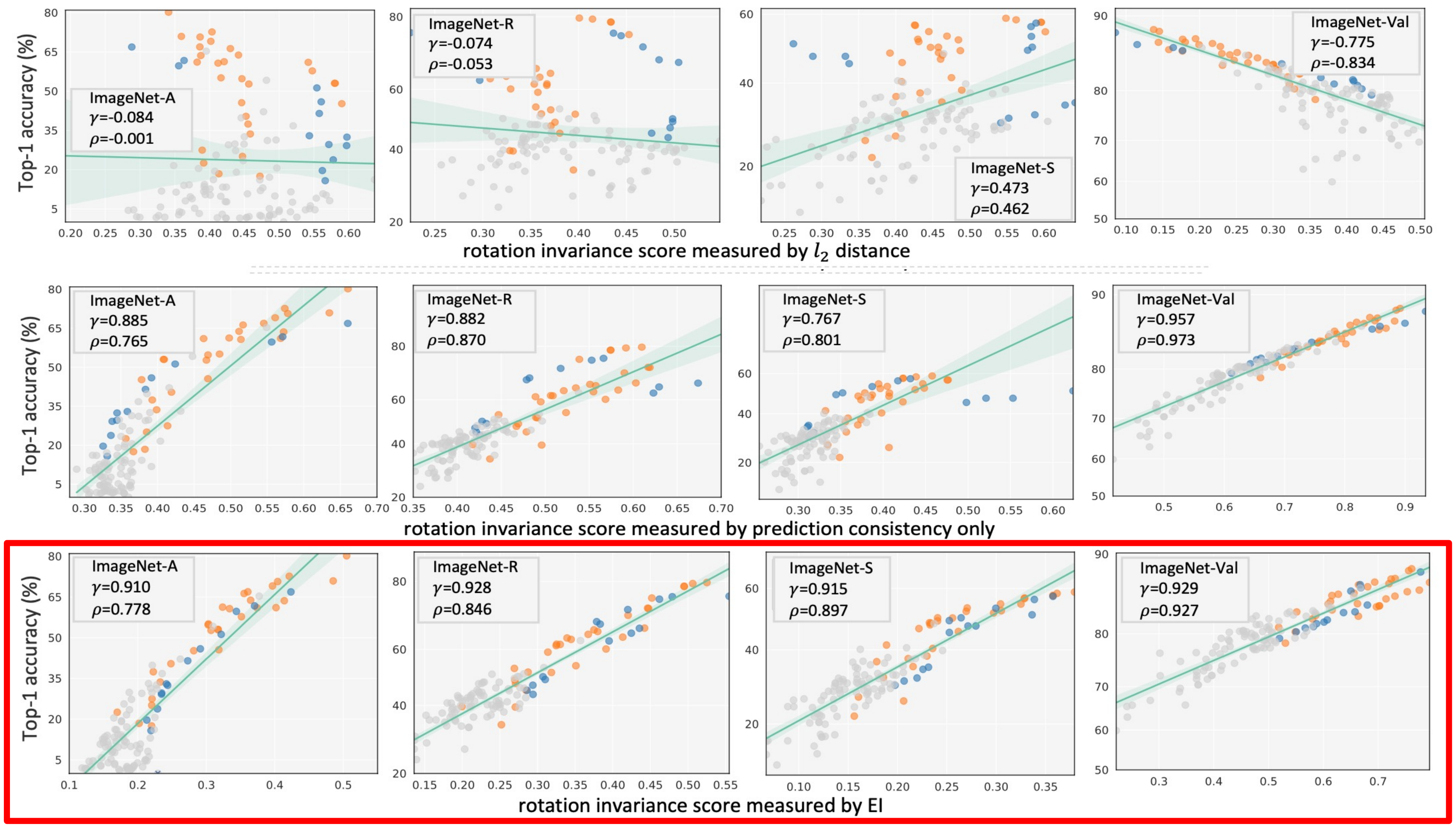
Training with more data allows models to achieve relatively **high accuracy and invariance** on three test sets

CIFAR-10 Setup

3 types of data distribution; 90 CIFAR-10 models



EI Gives Stronger Correlation Strength Than Other Measures



Summary

- Effective invariance (EI) to more reasonably measure invariance
- Classification accuracy and EI of various models has a strong linear relationship on both ID and OOD datasets

Thank you!

For more information,
please refer to
https://weijiandeng.xyz