# Learning under Distribution Shifts

16 June 2025

Myeongho Jeon

EPFL

1. How AI Systems are Trained and Make Predictions

2. Distribution Shifts and the Challenges They Pose

3. Approaches for Distribution Shifts

EPFL

1. **How AI Systems are Trained and Make Predictions**

2. Distribution Shifts and the Challenges They Pose

3. Approaches for Distribution Shifts

EPFL

# How AI Systems are Trained and Make Predictions

## Training Dataset

$$X = \{x_1, x_2, \cdots, x_N\}$$

$$Y = \{y_1, y_2, \cdots, y_N\}$$

**Dog (O)**
**Cat (X)**

**Dog (X)**
**Cat (O)**

EPFL

# How AI Systems are Trained and Make Predictions

## Machine Learning

$$\begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix}$$

# How AI Systems are Trained and Make Predictions

## Machine Learning

$$\begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix}$$

**Compare**

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

# How AI Systems are Trained and Make Predictions

## Machine Learning

$$\begin{pmatrix} 0.3 \\ 0.7 \end{pmatrix}$$

**Compare**

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

EPFL

# How AI Systems are Trained and Make Predictions

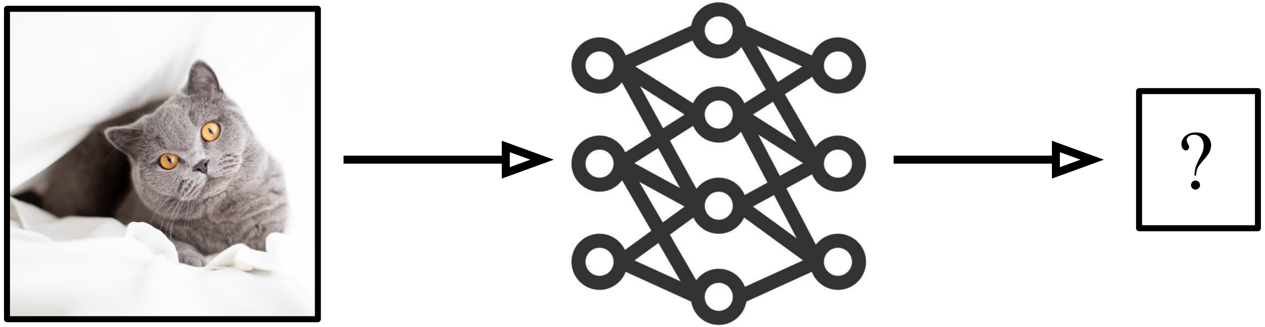
## Machine Learning

Train the model by **minimizing** the empirical risk

$$R(f) = \mathbb{E}_{(x,y)\sim p}\left[\mathcal{L}(y, f(x))\right]$$

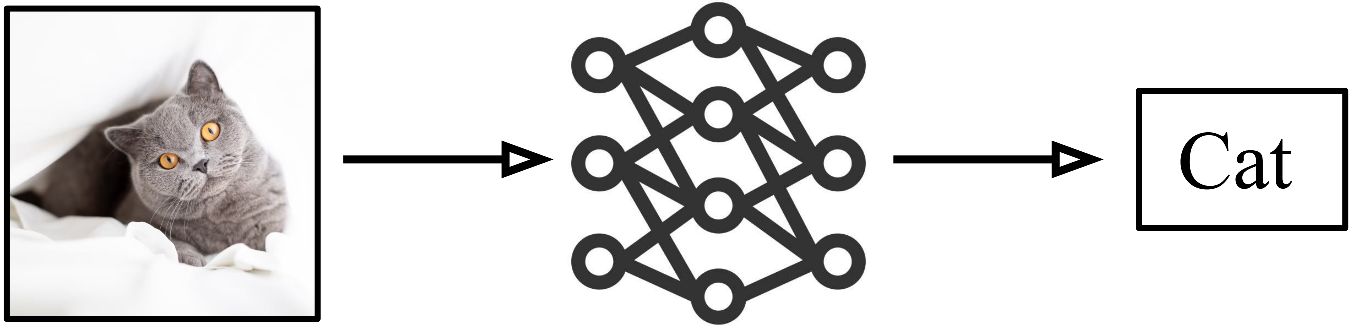
# How AI Systems are Trained and Make Predictions

## Machine Learning



# How AI Systems are Trained and Make Predictions

## Machine Learning




EPFL

1. How AI Systems are Trained and Make Predictions

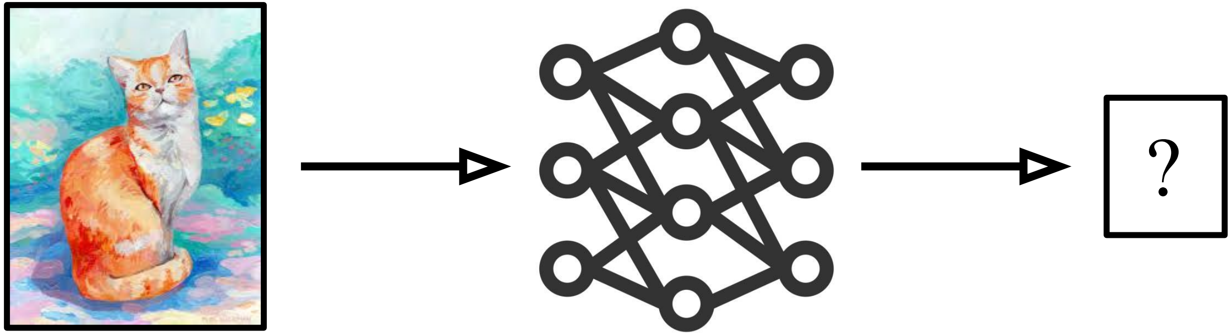
2. **Distribution Shifts and the Challenges They Pose**

3. Approaches for Distribution Shifts

EPFL

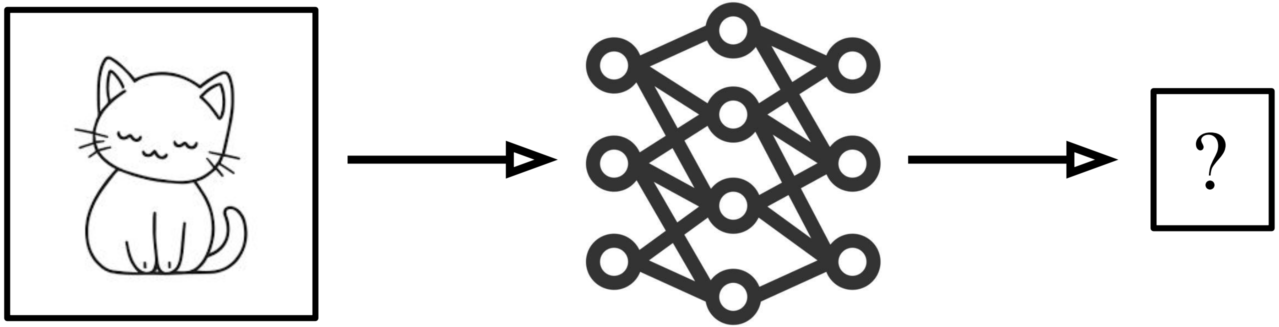
# Distribution Shifts and the Challenges They Pose

## Distribution Shifts



# Distribution Shifts and the Challenges They Pose

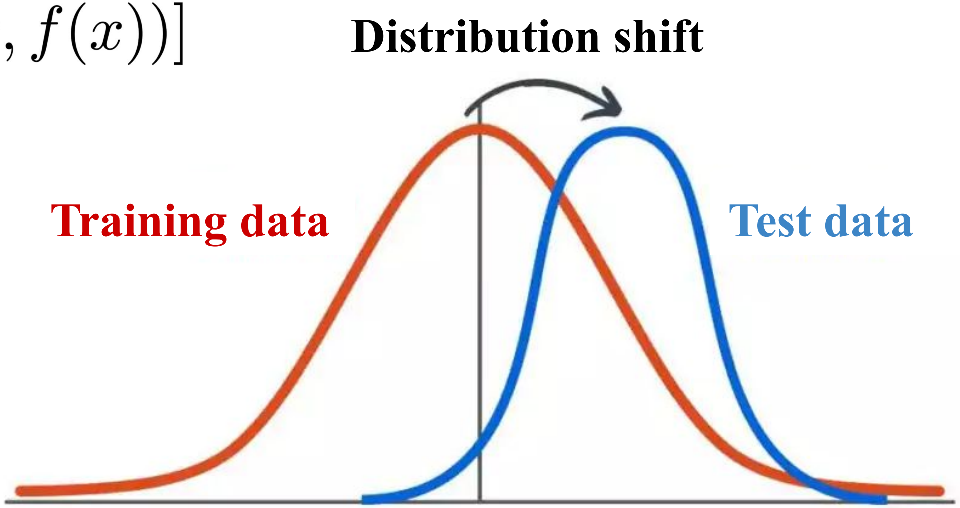
## Distribution Shifts



# Distribution Shifts and the Challenges They Pose

## Distribution Shifts

$$R(f) = \mathbb{E}_{(x,y)\sim p}\left[\mathcal{L}(y, f(x))\right]$$



# Distribution Shifts and the Challenges They Pose

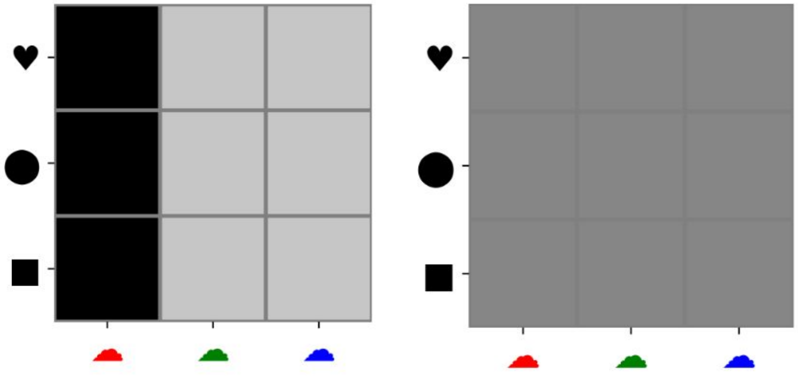
## Distribution Shifts

$$|R_{\text{train}}(f) - R_{\text{test}}(f)| = \left|\mathbb{E}_{(x,y)\sim p_{\text{train}}}\left[\mathcal{L}(y, f(x))\right] - \mathbb{E}_{(x,y)\sim p_{\text{test}}}\left[\mathcal{L}(y, f(x))\right]\right| > \epsilon.$$

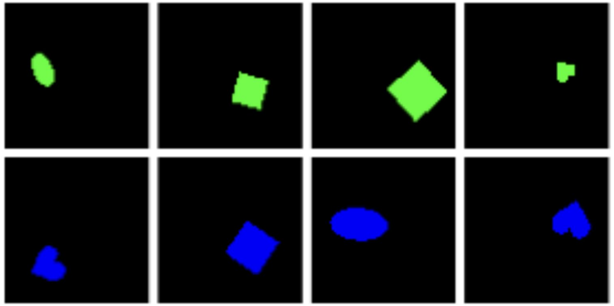
# Distribution Shifts and the Challenges They Pose

## Distribution Shift Types

### 1. Domain Shift



Training Data Test Data



Wiles, Olivia, et al. "A Fine-Grained Analysis on Distribution Shift." ICLR 2022.

# Distribution Shifts and the Challenges They Pose

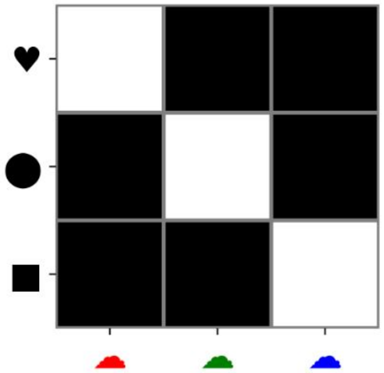
## Distribution Shift Types

### 1. Domain Shift

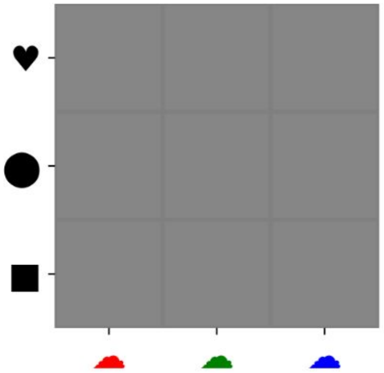
# Distribution Shifts and the Challenges They Pose

## Distribution Shift Types

### 2. Spurious Correlation



Training Data

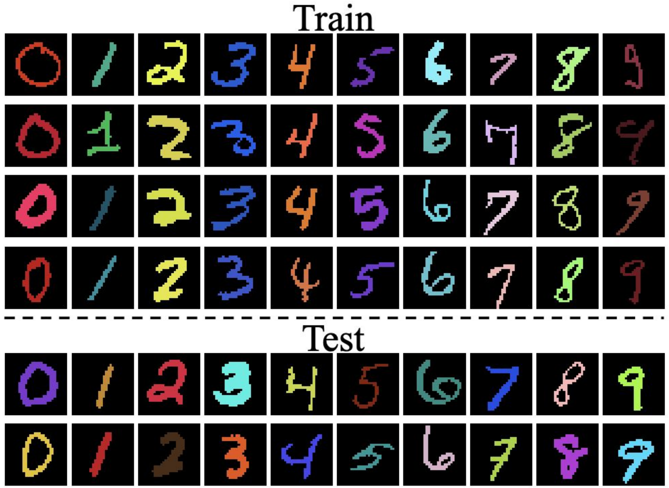


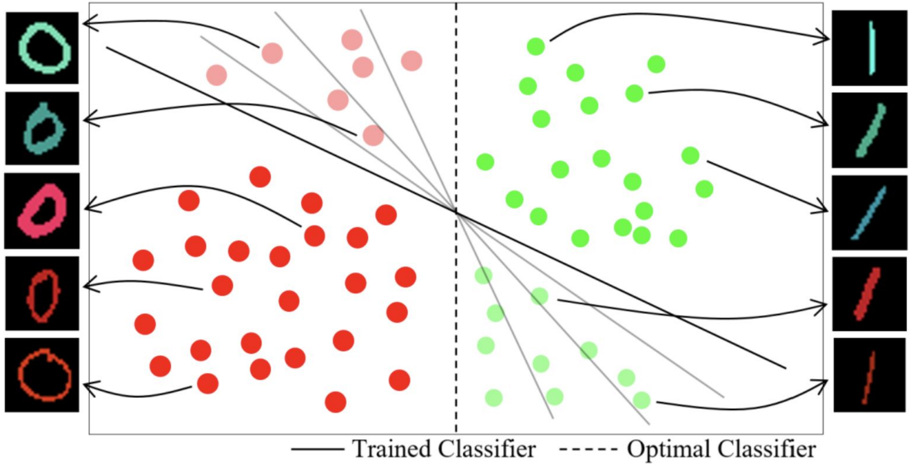
Test Data

# Distribution Shifts and the Challenges They Pose

## Distribution Shift Types

### 2. Spurious Correlation




Kim, Byungju, et al. "Learning not to learn: Training deep neural networks with biased data." CVPR 2019.

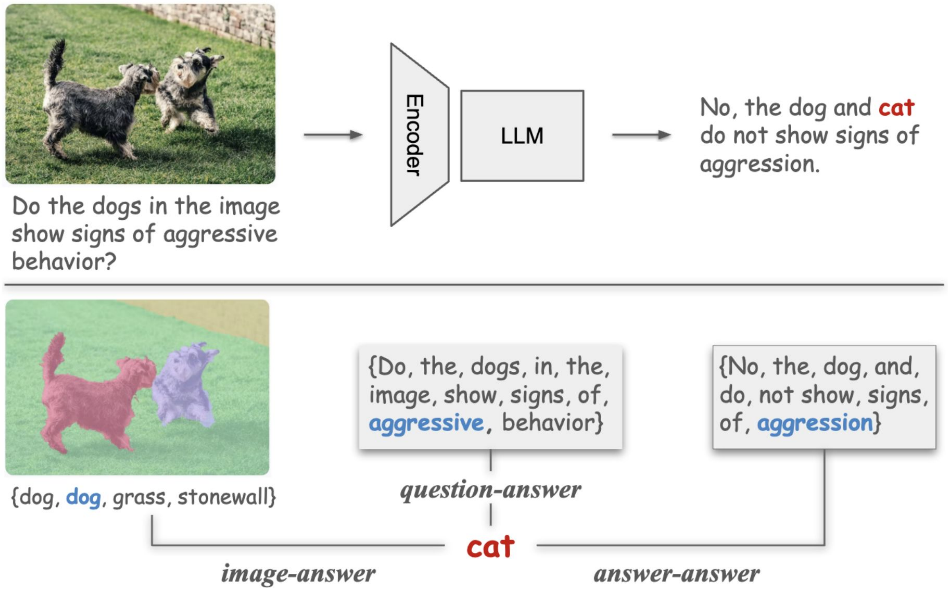
# Distribution Shifts and the Challenges They Pose

## Distribution Shift Types

### 2. Spurious Correlation

# Distribution Shifts and the Challenges They Pose

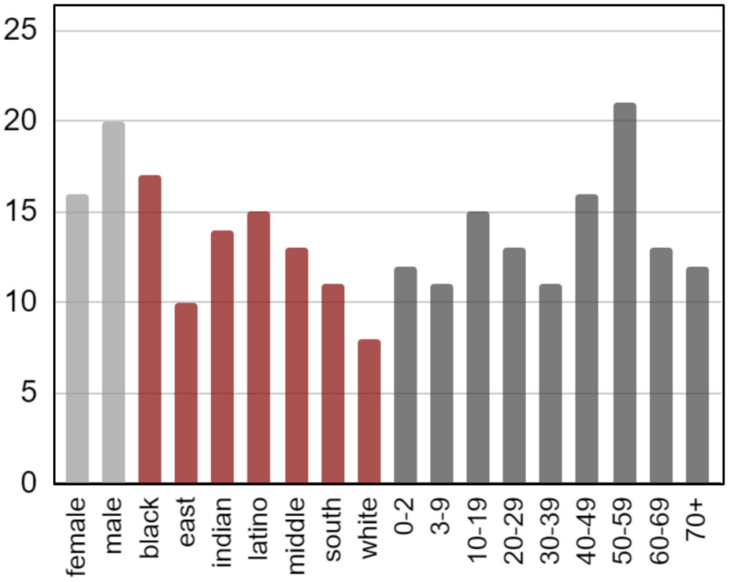
## Distribution Shift in Large Vision Language Model




EPFL

# Distribution Shifts and the Challenges They Pose

## Distribution Shift in Large Vision Language Model





EPFL

1. How AI Systems are Trained and Make Predictions

2. Distribution Shifts and the Challenges They Pose

3. **Approaches for Distribution Shifts**

# Approaches for Distribution Shifts

1. Benchmarks to evaluate distribution shifts

2. Model to address domain shift

3. Model to address spurious correlation

4. Zero-shot inference with foundation model (e.g., ChatGPT)

EPFL

# Approaches for Distribution Shifts

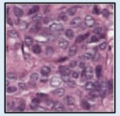
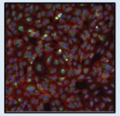
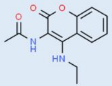
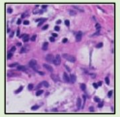
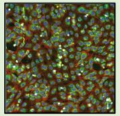
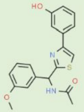
## Benchmark



Li, Da, et al. "Deeper, broader and artier domain generalization." ICCV 2017.

# Approaches for Distribution Shifts

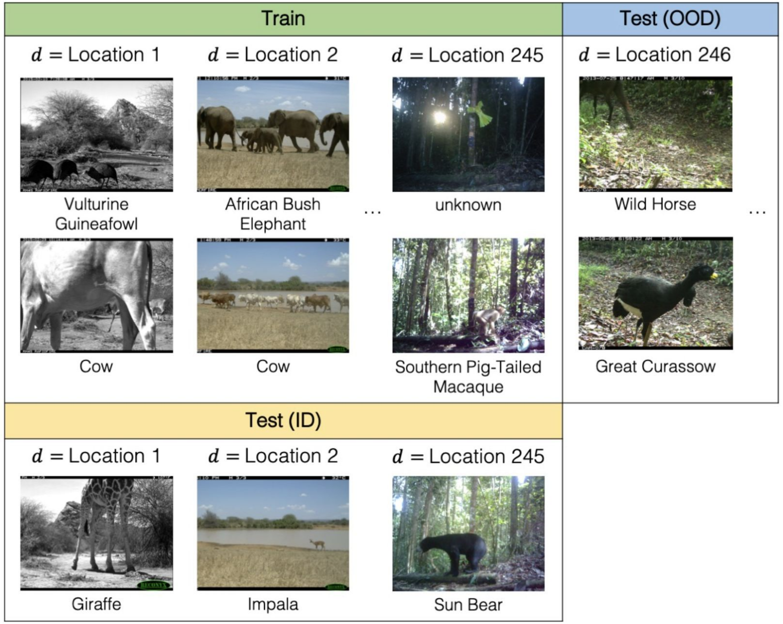
## Benchmark

| | Domain generalization | | | | | Subpopulation shift | Domain generalization + subpopulation shift | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Dataset | iWildCam | Camelyon17 | RxRx1 | OGB-MolPCBA | GlobalWheat | CivilComments | FMoW | PovertyMap | Amazon | Py150 |
| Input (x) | camera trap photo | tissue slide | cell image | molecular graph | wheat image | online comment | satellite image | satellite image | product review | code |
| Prediction (y) | animal species | tumor | perturbed gene | bioassays | wheat head bbox | toxicity | land use | asset wealth | sentiment | autocomplete |
| Domain (d) | camera | hospital | batch | scaffold | location, time | demographic | time, region | country, rural-urban | user | git repository |
| # domains | 323 | 5 | 51 | 120,084 | 47 | 16 | 16 x 5 | 23 x 2 | 2,586 | 8,421 |
| # examples | 203,029 | 455,954 | 125,510 | 437,929 | 6,515 | 448,000 | 523,846 | 19,669 | 539,502 | 150,000 |
| Train example |  |  |  |  |  | What do Black and LGBT people have to do with bicycle licensing? |  |  | Overall a solid package that has a good quality of construction for the price. | `import numpy as np`<br>`…`<br>`norm=np.___` |
| Test example |  |  |  |  |  | As a Christian, I will not be patronizing any of those businesses. |  |  | I *loved* my French press, it's so perfect and came with all this fun stuff! | `import subprocess as sp`<br>`p=sp.Popen()`<br>`stdout=p.___` |
| Adapted from | Beery et al. 2020 | Bandi et al. 2018 | Taylor et al. 2019 | Hu et al. 2020 | David et al. 2021 | Borkan et al. 2019 | Christie et al. 2018 | Yeh et al. 2020 | Ni et al. 2019 | Raychev et al. 2016 |

Koh, Pang Wei, et al. "Wilds: A benchmark of in-the-wild distribution shifts." ICML 2021.

# Approaches for Distribution Shifts

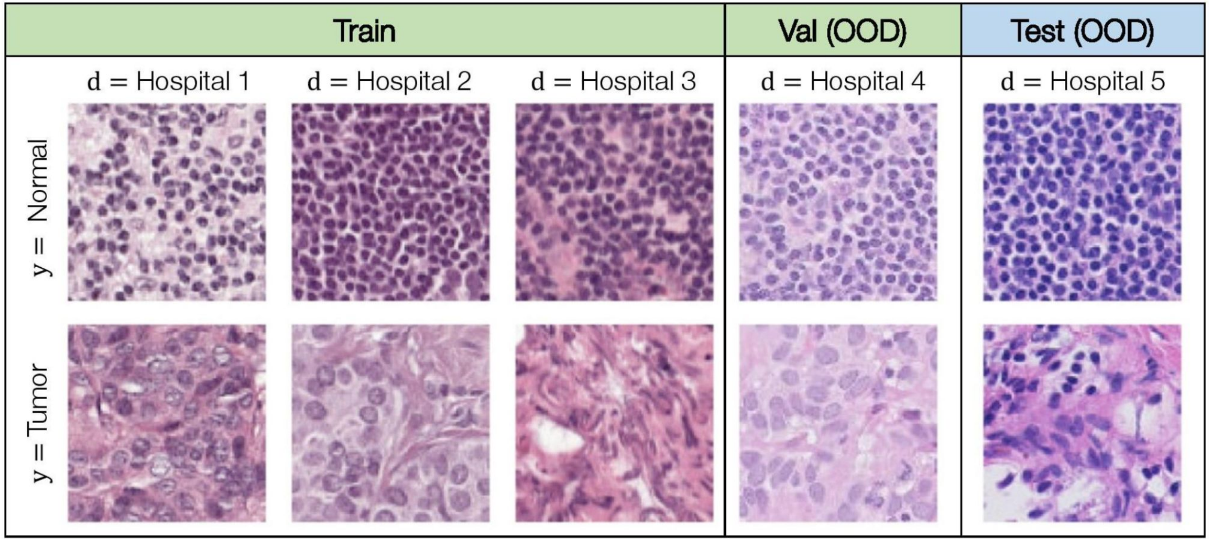
## Benchmark




Koh, Pang Wei, et al. "Wilds: A benchmark of in-the-wild distribution shifts." ICML 2021.

# Approaches for Distribution Shifts

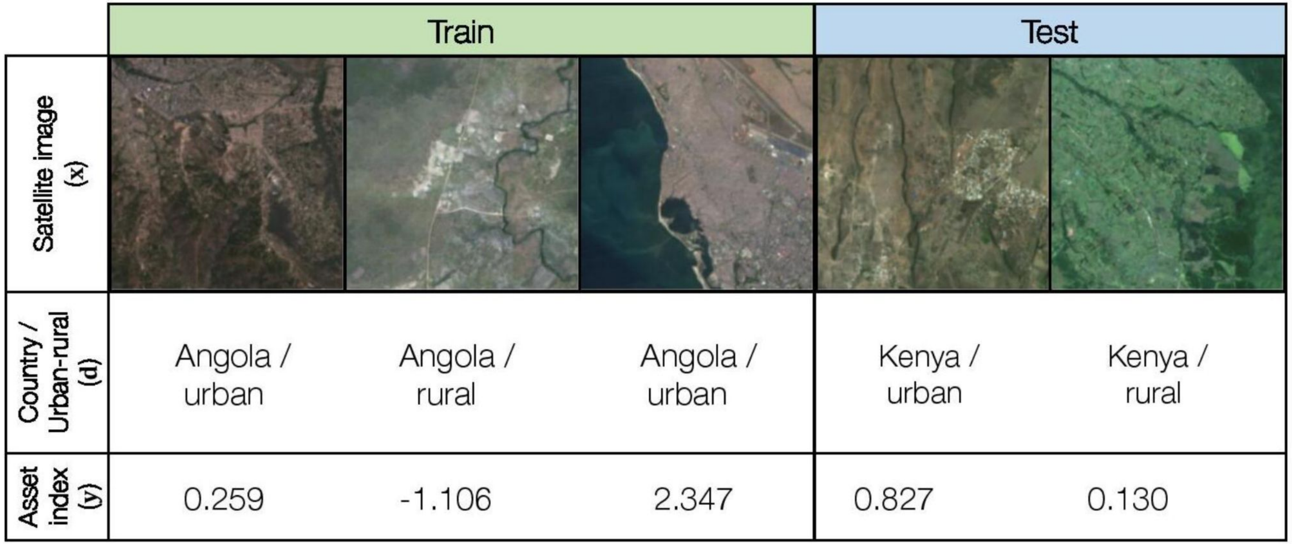
## Benchmark




Koh, Pang Wei, et al. "Wilds: A benchmark of in-the-wild distribution shifts." ICML 2021.

# Approaches for Distribution Shifts

## Benchmark




Koh, Pang Wei, et al. "Wilds: A benchmark of in-the-wild distribution shifts." ICML 2021.

# Approaches for Distribution Shifts

## Benchmark

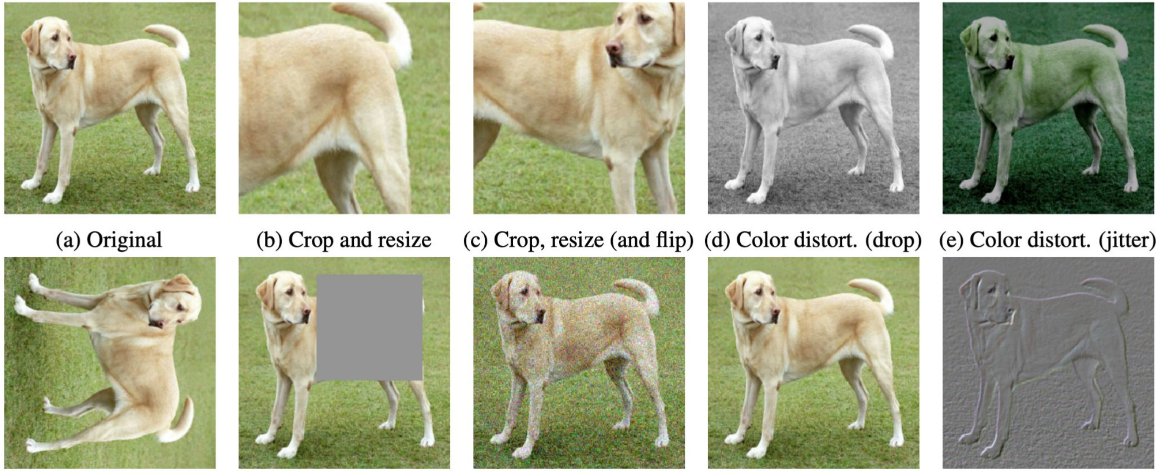
| | Repository ID ($d$) | Source code context ($x$) | Next tokens ($y$) |
|---|---|---|---|
| Train | Repository 1 | `... from easyrec.gateway import EasyRec <EOL> gateway = EasyRec('tenant','key') <EOL> item_type = gateway. ____` | `get_item_type` |
| | | `... response = gateway.get_other_users() <EOL> get_params = HTTPretty. ____` | `last_request` |
| | Repository 2 | `import numpy as np ... <EOL> if np.linalg.norm(target - prev_target) > far_threshold: <EOL> norm = np. ____` | `linalg` |
| | | `... new_trans = np.zeros((n_beats + max_beats, n_beats) <EOL> new_trans[:n_beats,:n_beats] = np. ____` | `max` |
| | ⋮ | | |
| Test | Repository 6,001 | `... if e.errno == errno.ENOENT: <EOL> continue <EOL> p = subprocess.Popen () <EOL> stdout = p. ____` | `communicate` |
| | | `... command = shlex.split(command) <EOL> command = map(str, command) <EOL> env = os. ____` | `environ` |
| | ⋮ | | |

Koh, Pang Wei, et al. "Wilds: A benchmark of in-the-wild distribution shifts." ICML 2021.

# Approaches for Domain Shifts

## Data Augmentation

"Let's help the model become more familiar with a wider variety of data"



(a) Original (b) Crop and resize (c) Crop, resize (and flip) (d) Color distort. (drop) (e) Color distort. (jitter)

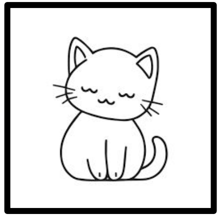
(f) Rotate $\{90°, 180°, 270°\}$ (g) Cutout (h) Gaussian noise (i) Gaussian blur (j) Sobel filtering

# Approaches for Domain Shifts

## Data Augmentation

“Let’s help the model become more familiar with a wider variety of data”

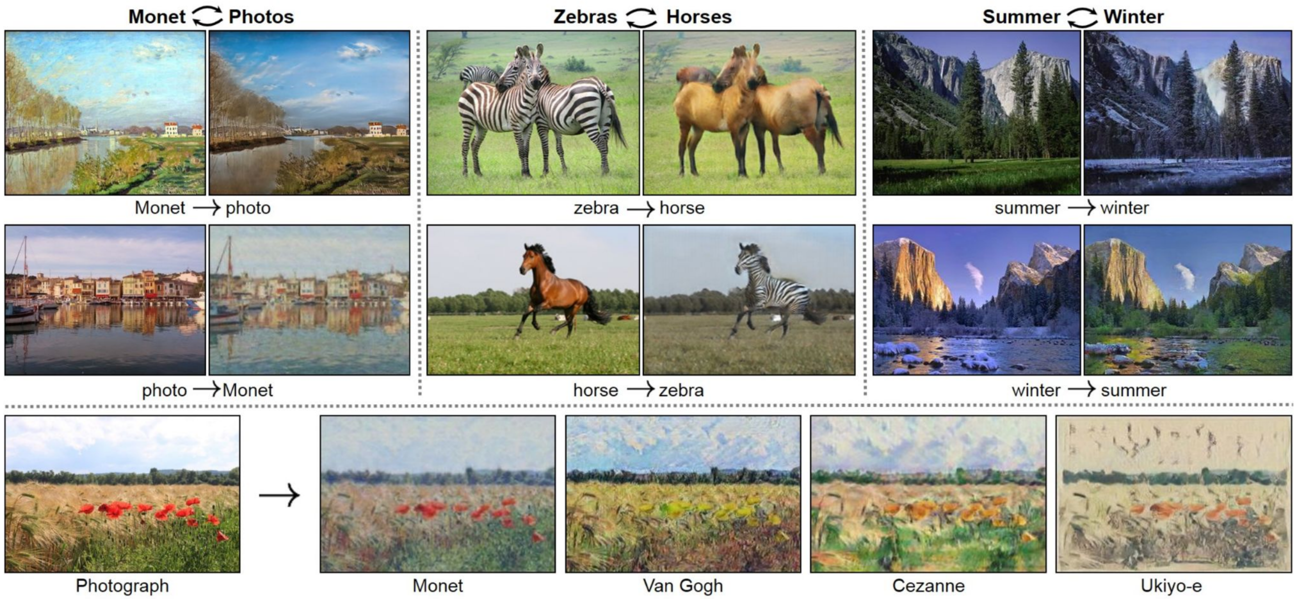
(a) Original (b) Crop and resize (c) Crop, resize (and flip) (d) Color distort. (drop) (e) Color distort. (jitter)



(f) Rotate $\{90°, 180°, 270°\}$ (g) Cutout (h) Gaussian noise (i) Gaussian blur (j) Sobel filtering

# Approaches for Domain Shifts

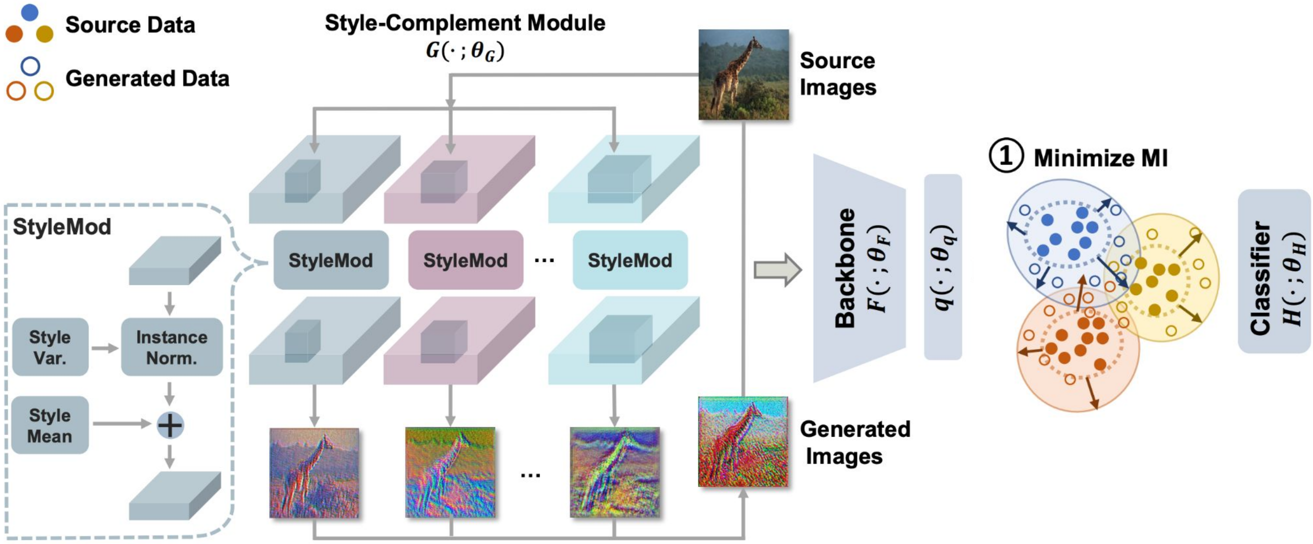
## Model-based Data Augmentation




Gatys, Leon A., et al. "Image style transfer using convolutional neural networks." CVPR 2016.

# Approaches for Domain Shifts

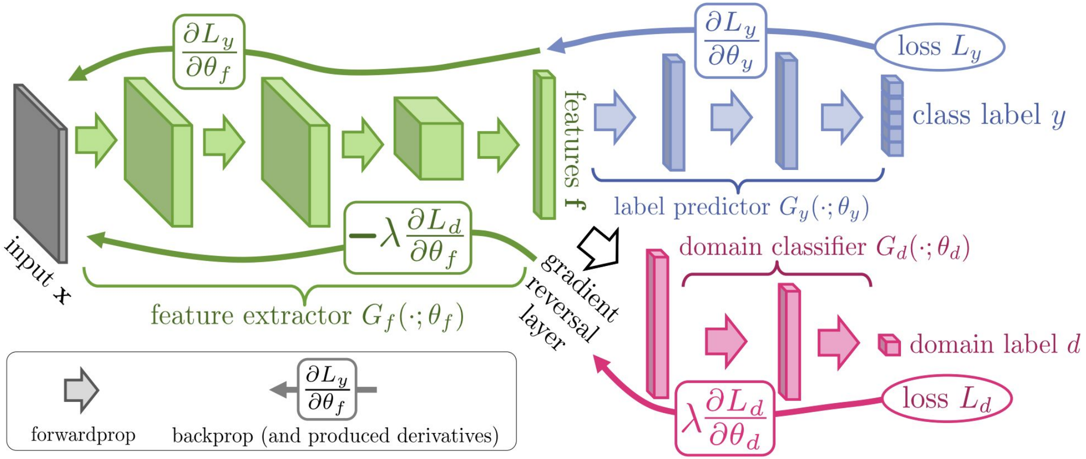
## Model-based Data Augmentation




Wang, Zijian, et al. "Learning to diversify for single domain generalization." ICCV 2021.

# Approaches for Domain Shifts

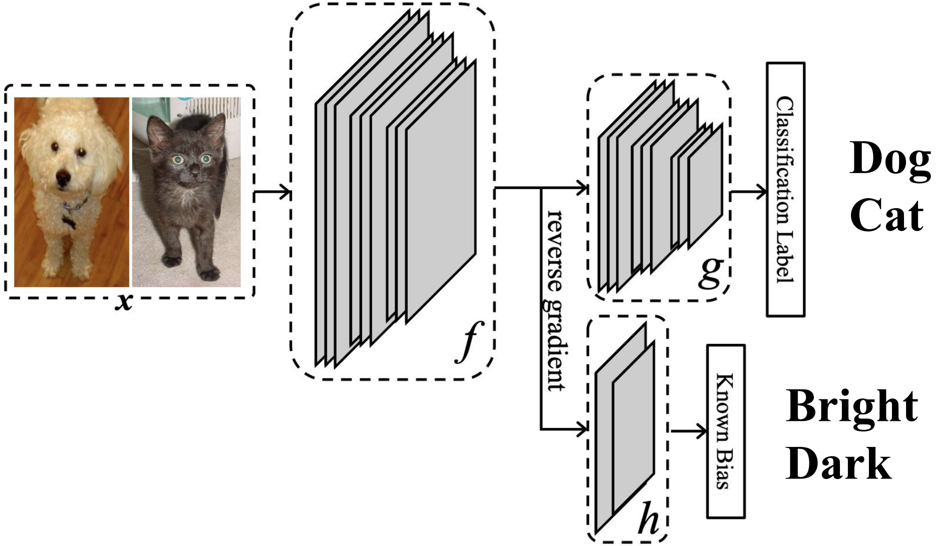
## Domain Adversarial Neural Network




Ganin, Yaroslav, et al. "Domain-adversarial training of neural networks." JMLR 2016.

# Approaches for Spurious Correlations

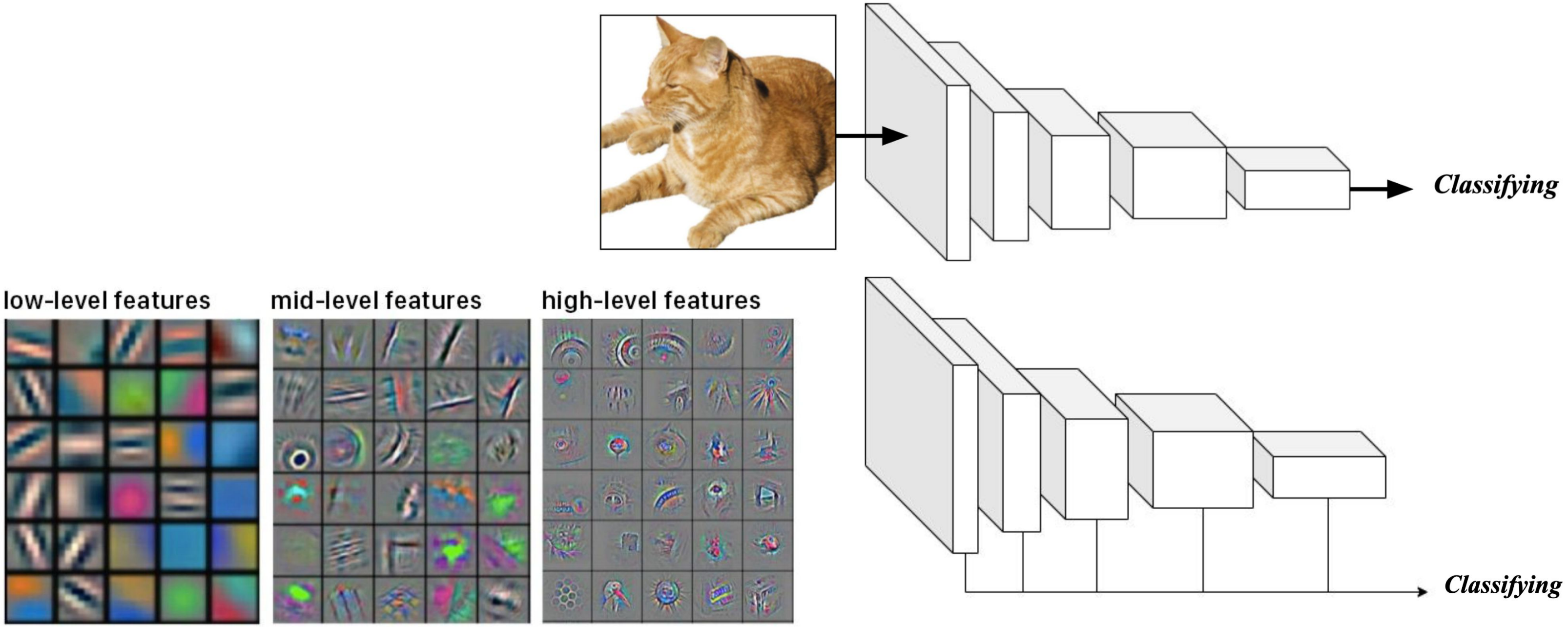
## Learning not to Learn




Kim, Byungju, et al. "Learning not to learn: Training deep neural networks with biased data." CVPR 2019.

# Approaches for Spurious Correlations

## Unbiasing Network



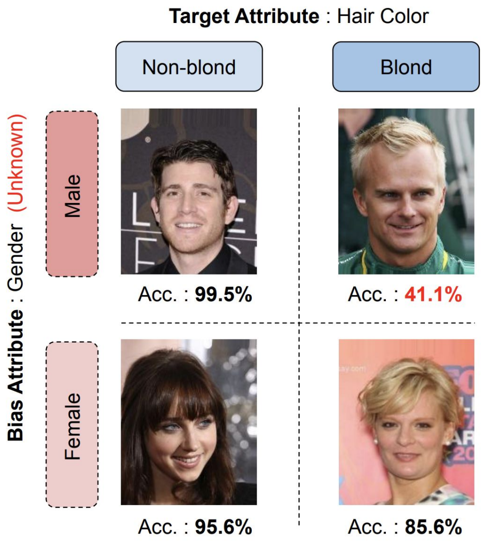

Jeon, Myeongho, et al. "A conservative approach for unbiased learning on unknown biases." CVPR 2022.

# Approaches for Spurious Correlations

## Weighted Resampling

"Let's assign higher weights to underrepresented data during training"

| | |
|---|---|
| 100 | 10 |
| 95 | 90 |




$$w_g \, \ell(\theta; (x, y))$$

| | |
|---|---|
| 0.1 | 1 |
| 0.15 | 0.2 |

Seo, Seonguk, et. al., "Unsupervised learning of debiased representations with pseudo-attributes." CVPR 2022.

# Approaches for Distribution Shifts

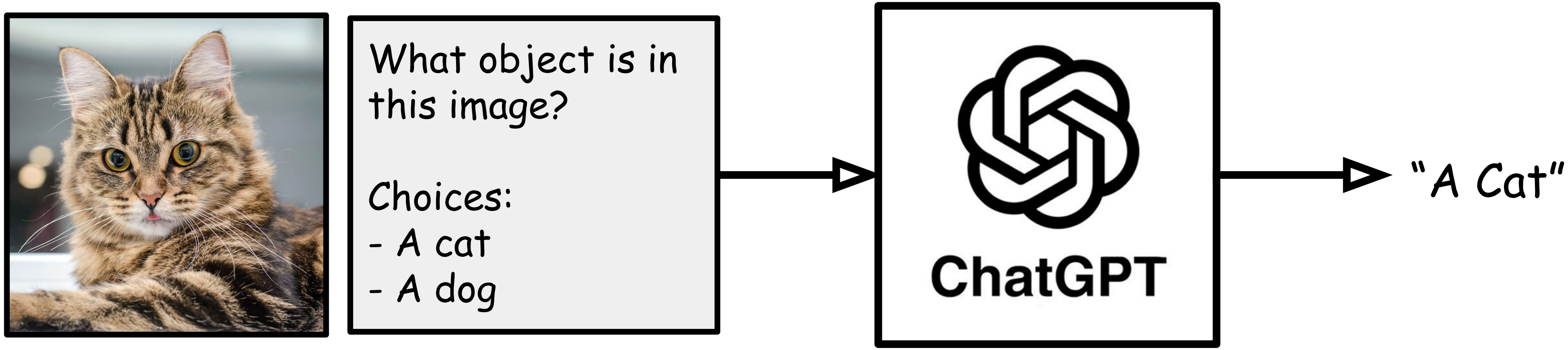
"Let's use a foundation model that has been trained on a large amount of data!"


ChatGPT

EPFL

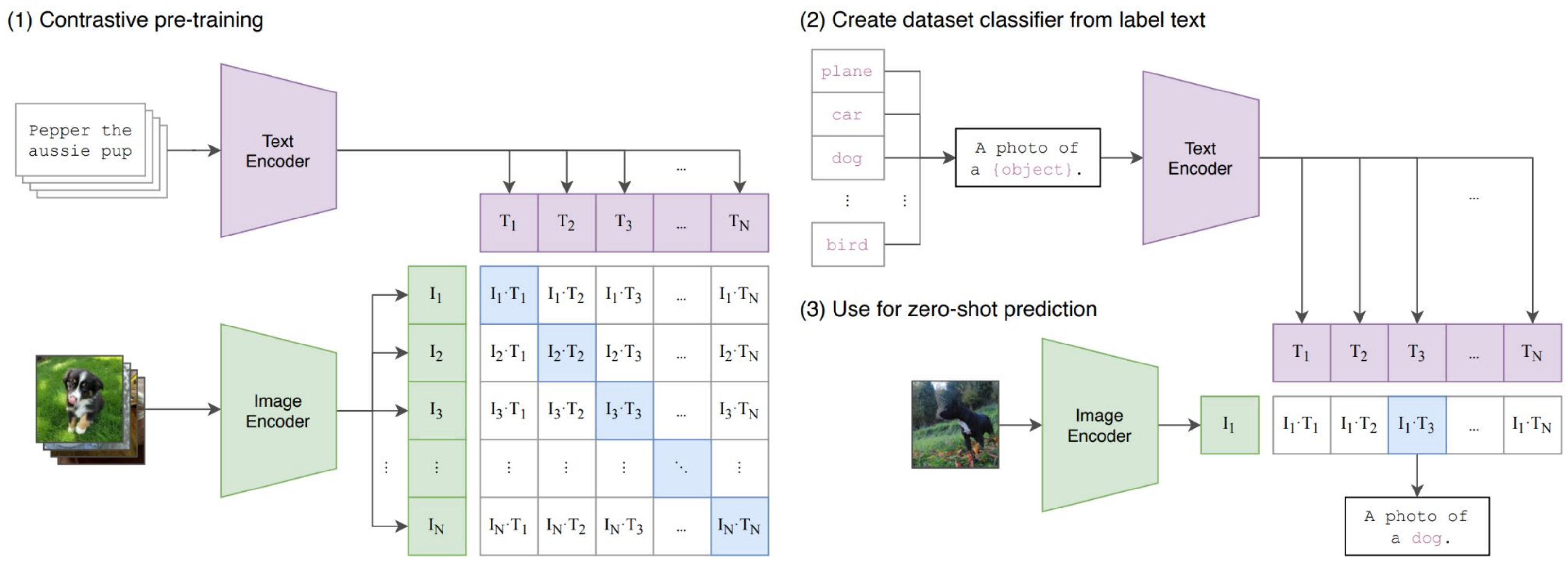
# Approaches for Distribution Shifts

## Zero-shot Inference with Foundation Model



# Approaches for Distribution Shifts

## Zero-shot Inference with Foundation Model




Radford, Alec, et al. "Learning transferable visual models from natural language supervision." ICML 2021.

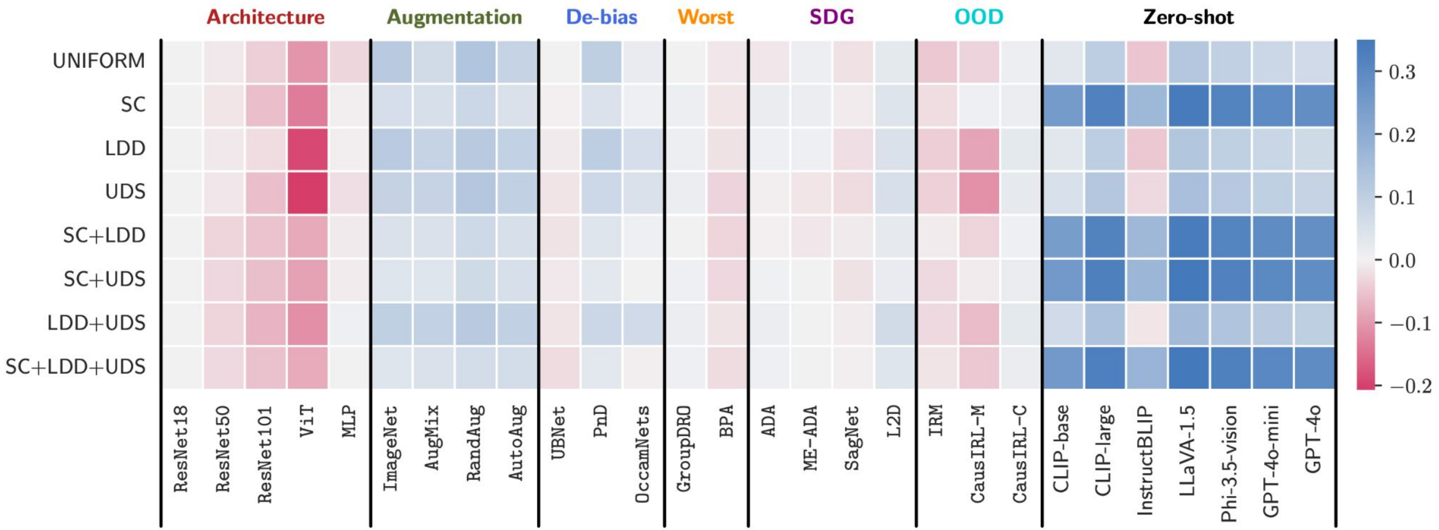
# Approaches for Distribution Shifts

## Comparison of approaches

| *Generalization Algorithms* | |
|---|---|
| *Architecture* | ResNet18, ResNet50, ResNet101 (He et al., 2016), ViT (Dosovitskiy et al., 2020), MLP (Vapnik, 1991). |
| *Heuristic augmentation* | Imagenet (He et al., 2016), AugMix (Hendrycks et al., 2019), RandAug (Cubuk et al., 2020), AutoAug (Cubuk et al., 2019). |
| *De-biasing* | UBNet (Jeon et al., 2022a), PnD (Li et al., 2023), OccamNets (Shrestha et al., 2022). |
| *Worst-case generalization* | GroupDRO (Sagawa et al., 2019), BPA (Seo et al., 2022). |
| *Single domain generalization* | ADA (Volpi et al., 2018), ME-ADA (Zhao et al., 2020), SagNet (Nam et al., 2021), L2D (Wang et al., 2021). |
| *Out-of-distribution generalization* | IRM (Arjovsky et al., 2019), CausIRL (Chevalley et al., 2022). |
| *Zero-shot inference with foundation model* | CLIP (Radford et al., 2021), InstructBLIP (Dai et al., 2024), LLaVA-1.5 (Liu et al., 2023), Phi-3.5-vision (Abdin et al., 2024), GPT-4o-mini, GPT-4o (OpenAI, 2024). |

# Approaches for Distribution Shifts

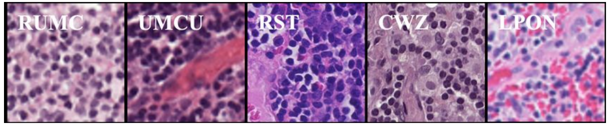
1. Simple data augmentation is more effective than state-of-the-art models.
2. Zero-shot inference with foundation model is very effective.

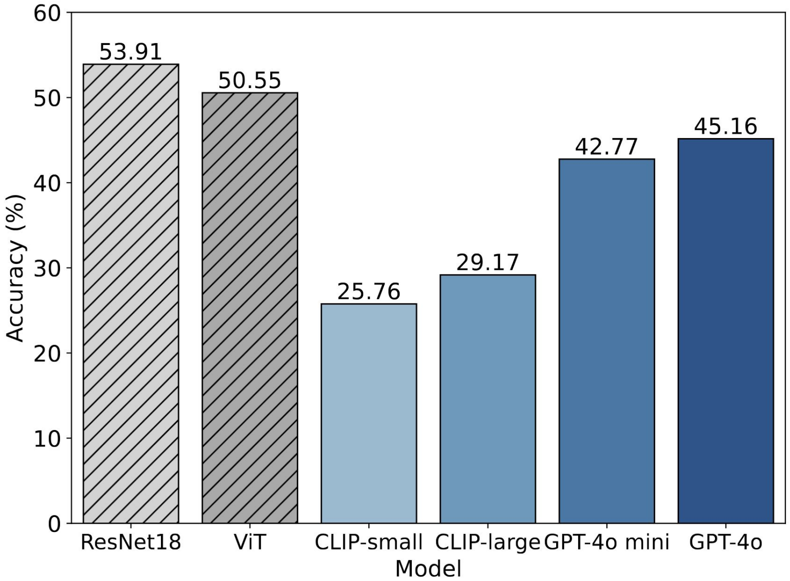



Jeon et al. "An Analysis of Model Robustness across Concurrent Distribution Shifts." TMLR 2025.

# Approaches for Distribution Shifts

## Distribution shifts in foundation model



# Reference


[1] Wiles, Olivia, et al. "A Fine-Grained Analysis on Distribution Shift." ICLR 2022.

[2] Zhang, Yue, et al. "Alleviating Hallucinations of Large Language Models through Induced Hallucinations." CoRR 2023.

[3] Jeon, Myeongho, et al. "An Analysis of Model Robustness across Concurrent Distribution Shifts." TMLR 2025.

[4] Wang, Zijian, et al. "Learning to diversify for single domain generalization." ICCV 2021.

[5] Koh, Pang Wei, et al. "Wilds: A benchmark of in-the-wild distribution shifts." ICML 2021.

[6] Kim, Byungju, et al. "Learning not to learn: Training deep neural networks with biased data." CVPR 2019.

[7] Jeon, Myeongho, et al. "A conservative approach for unbiased learning on unknown biases." CVPR 2022.

[8] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." ICML 2021.

[9] Li, Da, et al. "Deeper, broader and artier domain generalization." ICCV 2017.

[10] Ganin, Yaroslav, et al. "Domain-adversarial training of neural networks." JMLR 2016.

[11] Gatys, Leon A., et al.  "Image style transfer using convolutional neural networks." CVPR 2016.

[12] Seo, Seonguk, et. al., "Unsupervised learning of debiased representations with pseudo-attributes." CVPR 2022.

# Thank you!