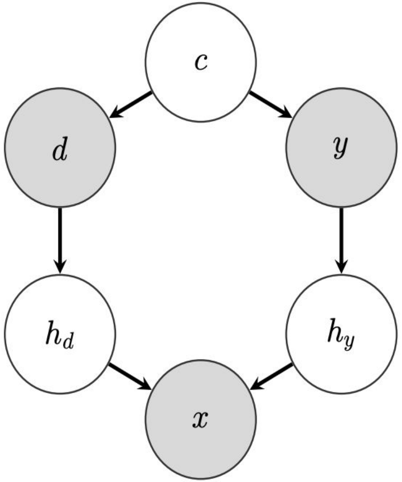
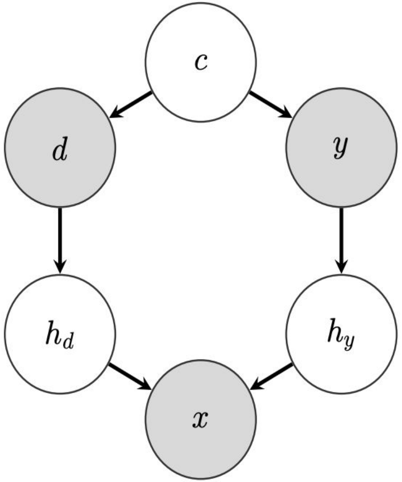
Selecting Data Augmentation for Simulating Interventions

Maximilian Ilse, Jakub M. Tomczak, Patrick Forré

Domain Generalization



Domain Generalization



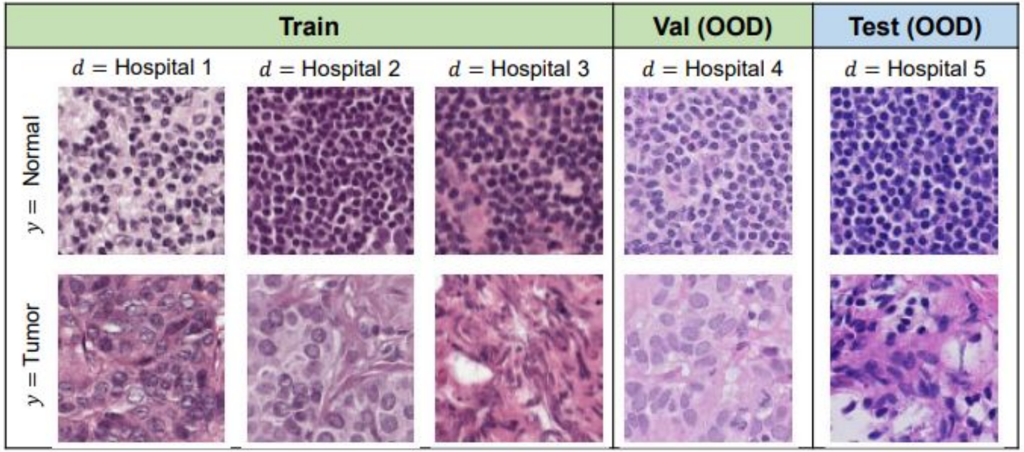
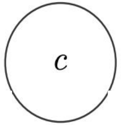


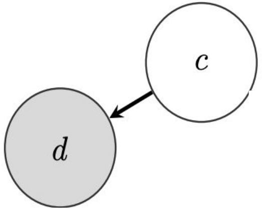
Image taken from Koh et al. - WILDS

Domain Generalization



c confounder: geographical location

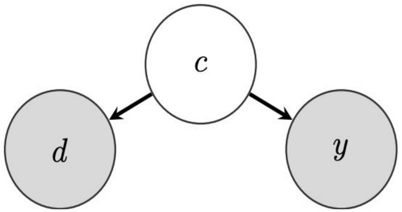
Domain Generalization



c confounder: geographical location

d domain: hospital

Domain Generalization

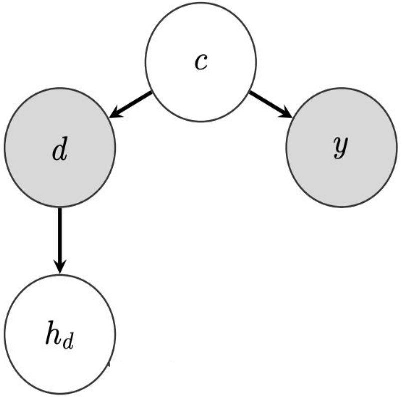


c confounder: geographical location

d domain: hospital

y label: disease

Domain Generalization



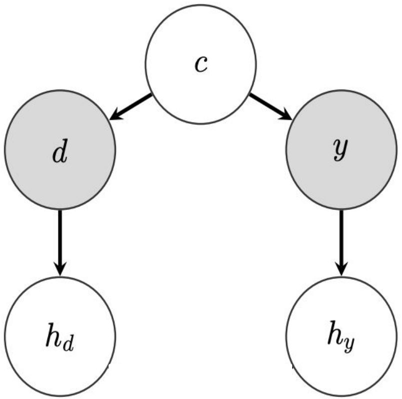
c confounder: geographical location

d domain: hospital

y label: disease

h_d high level features caused by d: color

Domain Generalization



c confounder: geographical location

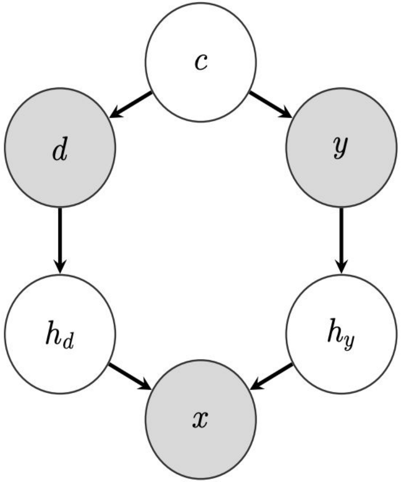
d domain: hospital

y label: disease

h_d high level features caused by d: color

h_y high level features caused by y: morphology

Domain Generalization



c confounder: geographical location

d domain: hospital

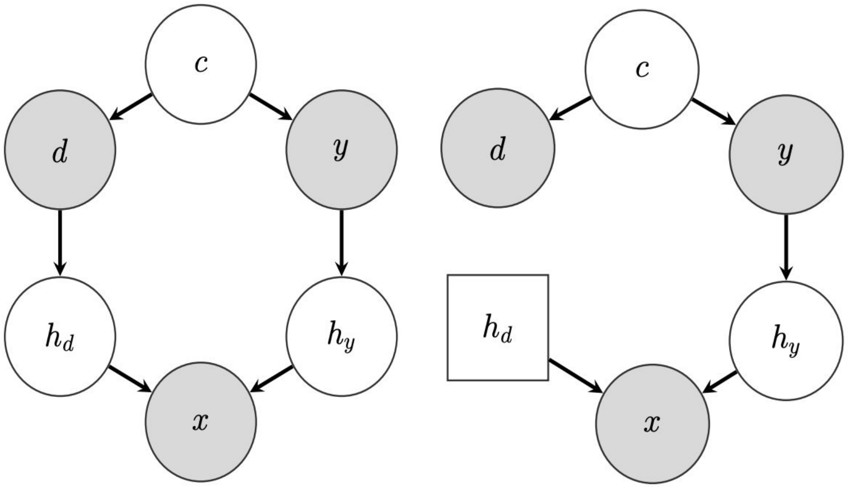
y label: disease

h_d high level features caused by d: color

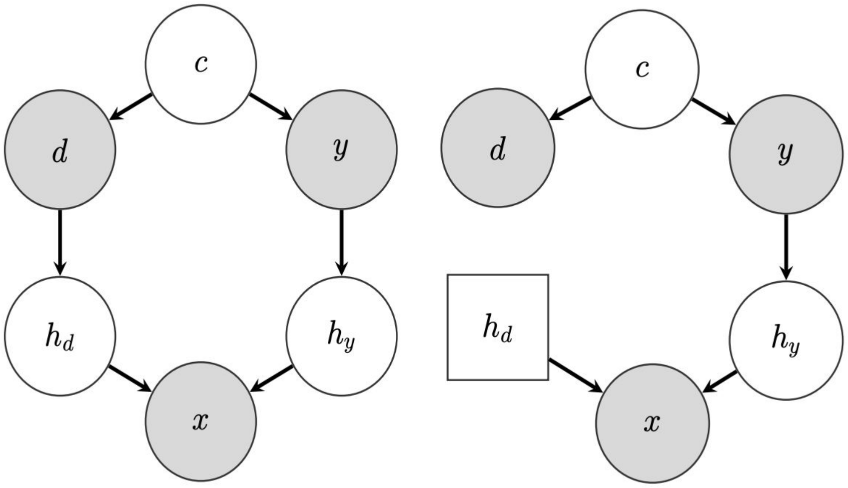
h_y high level features caused by y: morphology

x observation: image

Intervention



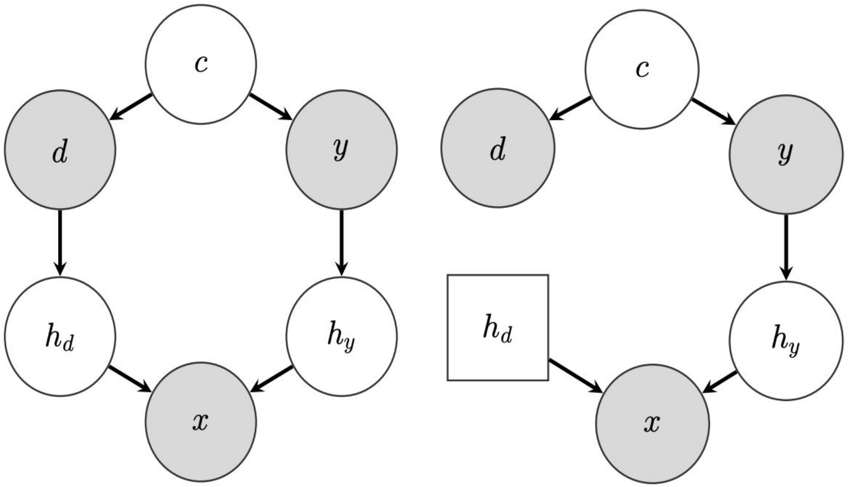
Intervention



Noise intervention:

$$\mathrm{do}(\mathbf{h}_d = \xi), \text{ where } \xi \sim N_\xi$$

Intervention

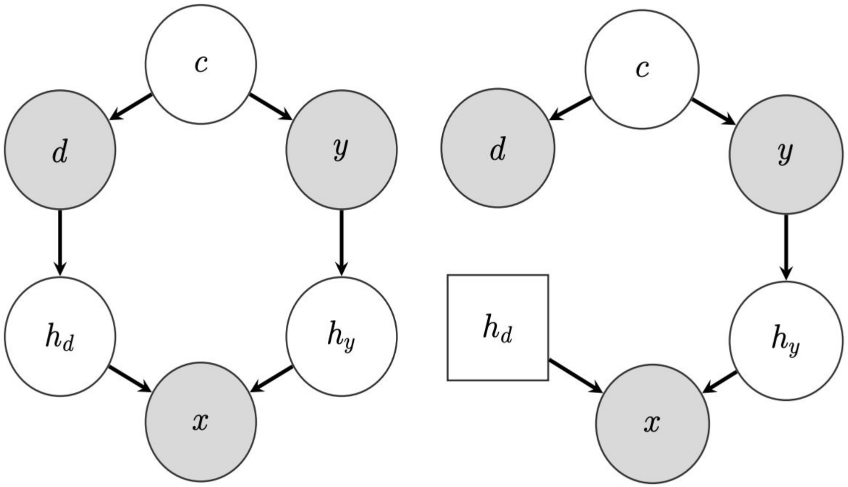


Noise intervention:

$$\mathrm{do}(\mathbf{h}_d = \xi), \text{ where } \xi \sim N_\xi$$

Removes spurious correlation

Intervention



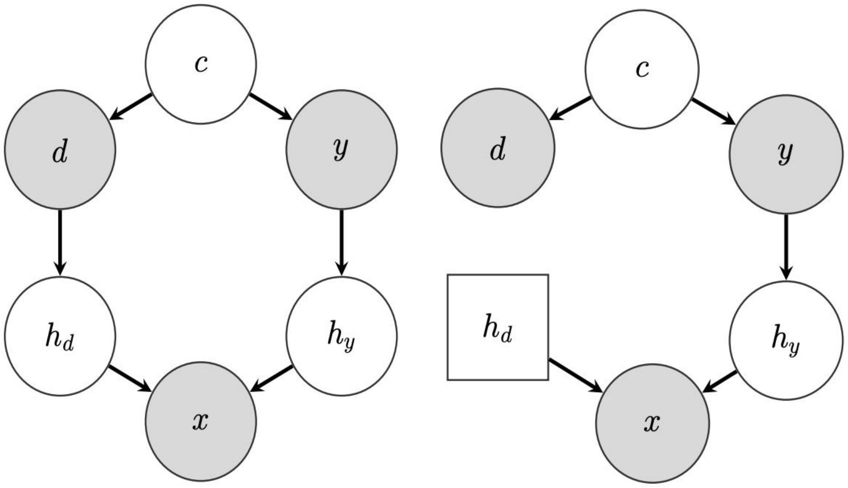
Noise intervention:

$$\mathrm{do}(\mathbf{h}_d = \xi), \text{ where } \xi \sim N_\xi$$

Removes spurious correlation

Interventions need to happen in the real world

Intervention



Noise intervention:

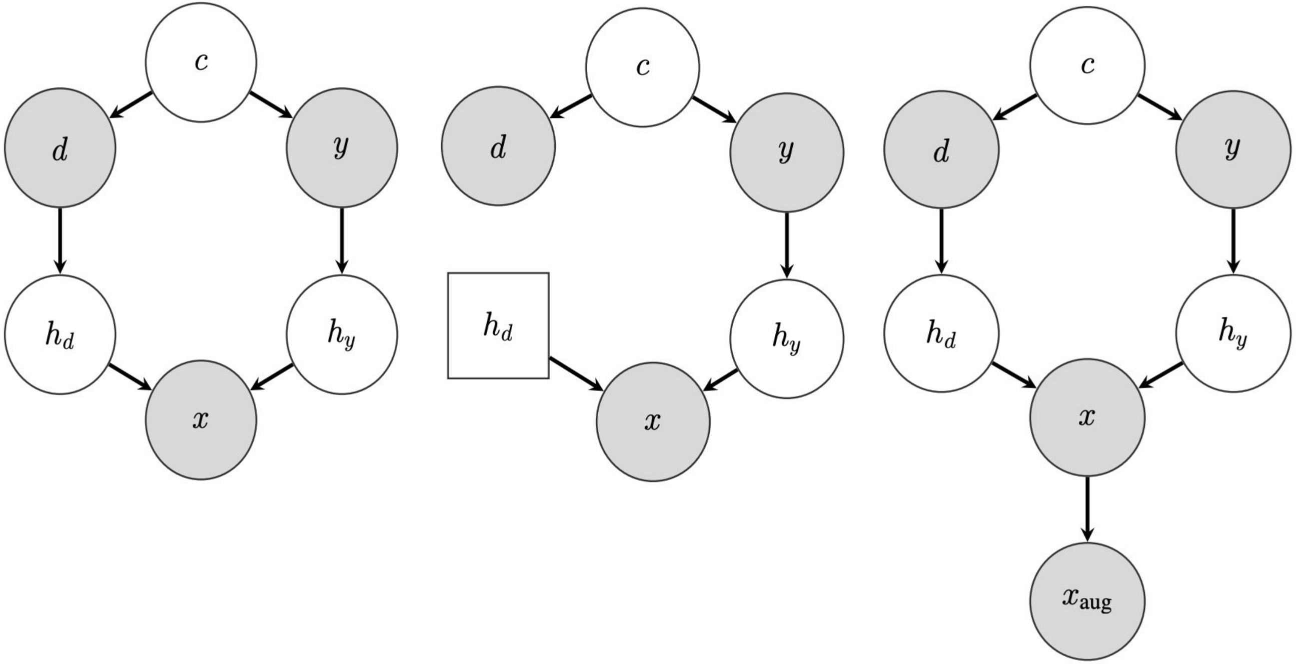
$$\mathrm{do}(\mathbf{h}_d = \xi), \text{ where } \xi \sim N_\xi$$

Removes spurious correlation

Interventions need to happen in the real world

Data Augmentation

Data augmentation



$$
\begin{tikzcd}
(h_d, h_y) \arrow[r, mapsto, "f_X"] \arrow[d, mapsto, "\mathrm{do}"'] & x \arrow[d, mapsto, "\mathrm{aug}"] \\
(\mathrm{do}(h_d), h_y) \arrow[r, mapsto, "f_X"] & x_{\mathrm{aug}}
\end{tikzcd}
$$

Figure 3: Intervention-augmentation equivariance expressed in a commutative diagram.

Select Data Augmentation (SDA)

Table 1: Results on Rotated MNIST results. Average accuracy for ten seeds.

Target	ERM	DANN	CDANN	SDA
0°	75.4	77.1	78.5	**96.1**
30°	93.4	94.2	94.9	**95.9**
60°	94.5	95.2	95.6	**95.7**
90°	79.6	83.0	84.0	**95.9**
Ave	85.7	87.4	88.3	**95.9**

Table 3: Results on Colored MNIST. Average accuracy ± standard deviation for ten seeds.

Acc	ERM	IRM	REx	SDA
Train	**87.4 ± 0.2**	70.8 ± 0.9	71.5 ± 1.0	72.1 ± 0.4
Test	17.1 ± 0.6	66.9 ± 2.5	68.7 ± 0.9	**74.1 ± 0.9**

Table 2: Results on PACS dataset. Average accuracy for five seeds.

Target	ERM	CDANN	L2G	GLCM	SSN	IRM	REx	MetaReg	JigSaw	SDA
A	63.3	62.7	66.2	66.8	64.1	67.1	67.0	69.8	67.6	**70.45**
C	63.1	69.7	66.9	69.7	66.8	68.5	68.0	70.4	**71.7**	68.49
P	87.7	78.7	88.0	87.9	90.2	89.4	89.7	**91.1**	89.0	88.35
S	54.1	64.5	59.0	56.3	60.1	57.8	59.8	59.3	65.2	**72.24**
Ave	67.1	68.9	70.0	70.2	70.3	70.7	71.1	72.6	73.4	**74.9**