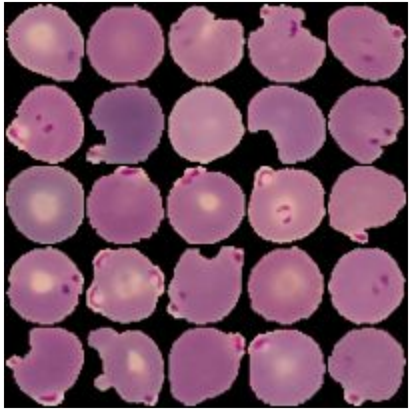
# DIVA:
# Domain Invariant Variational Autoencoders

In collaboration with Jakub Tomczak,
Christos Louizos and Max Welling

# Why do we care about domain generalization/invariance?

# Domain shift in medical imaging

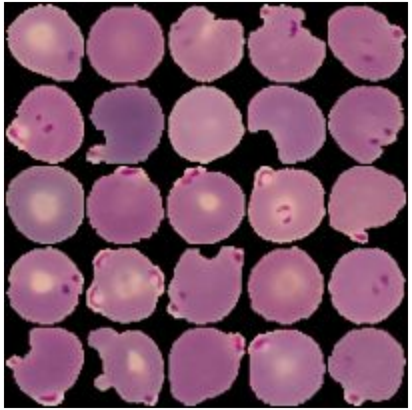
Patient 1



(Rajaraman et al., 2018)

# Domain shift in medical imaging

Patient 1

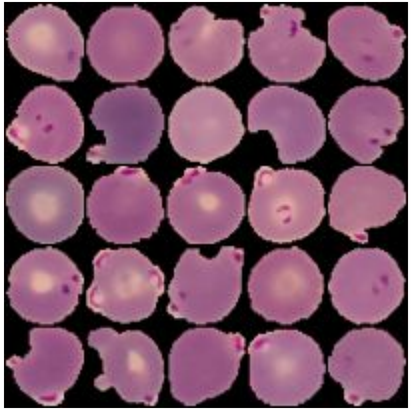


Malaria dataset

(Rajaraman et al., 2018)

# Domain shift in medical imaging

Patient 1

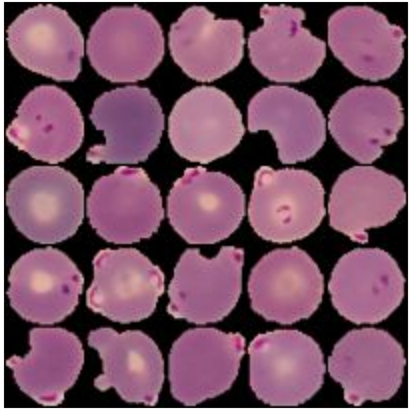


Malaria dataset
1 cell == 1 image

(Rajaraman et al., 2018)

# Domain shift in medical imaging

Patient 1

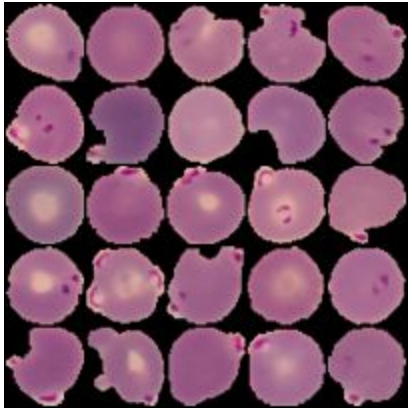


Malaria dataset
1 cell == 1 image
Task: infected vs. uninfected

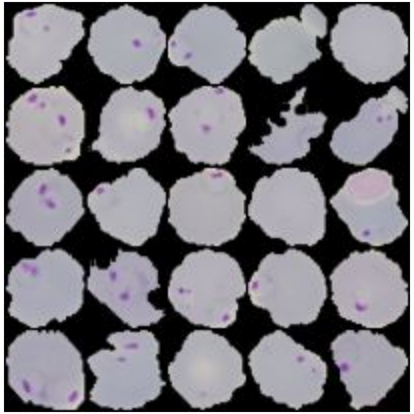
(Rajaraman et al., 2018)

# Domain shift in medical imaging

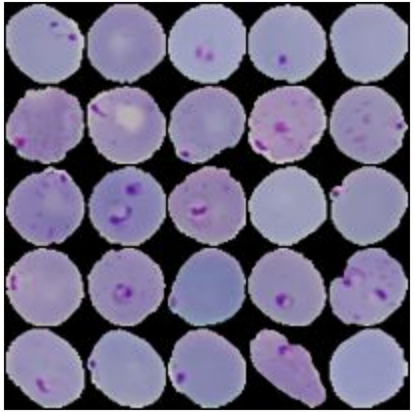
Patient 1



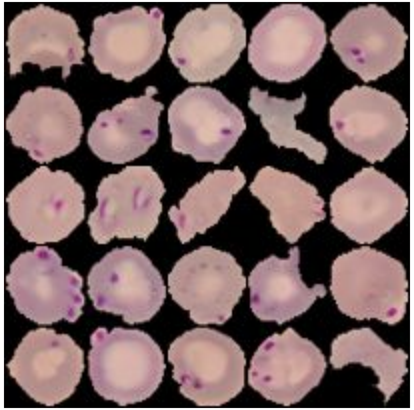
Patient 2



Patient 3



Patient 4

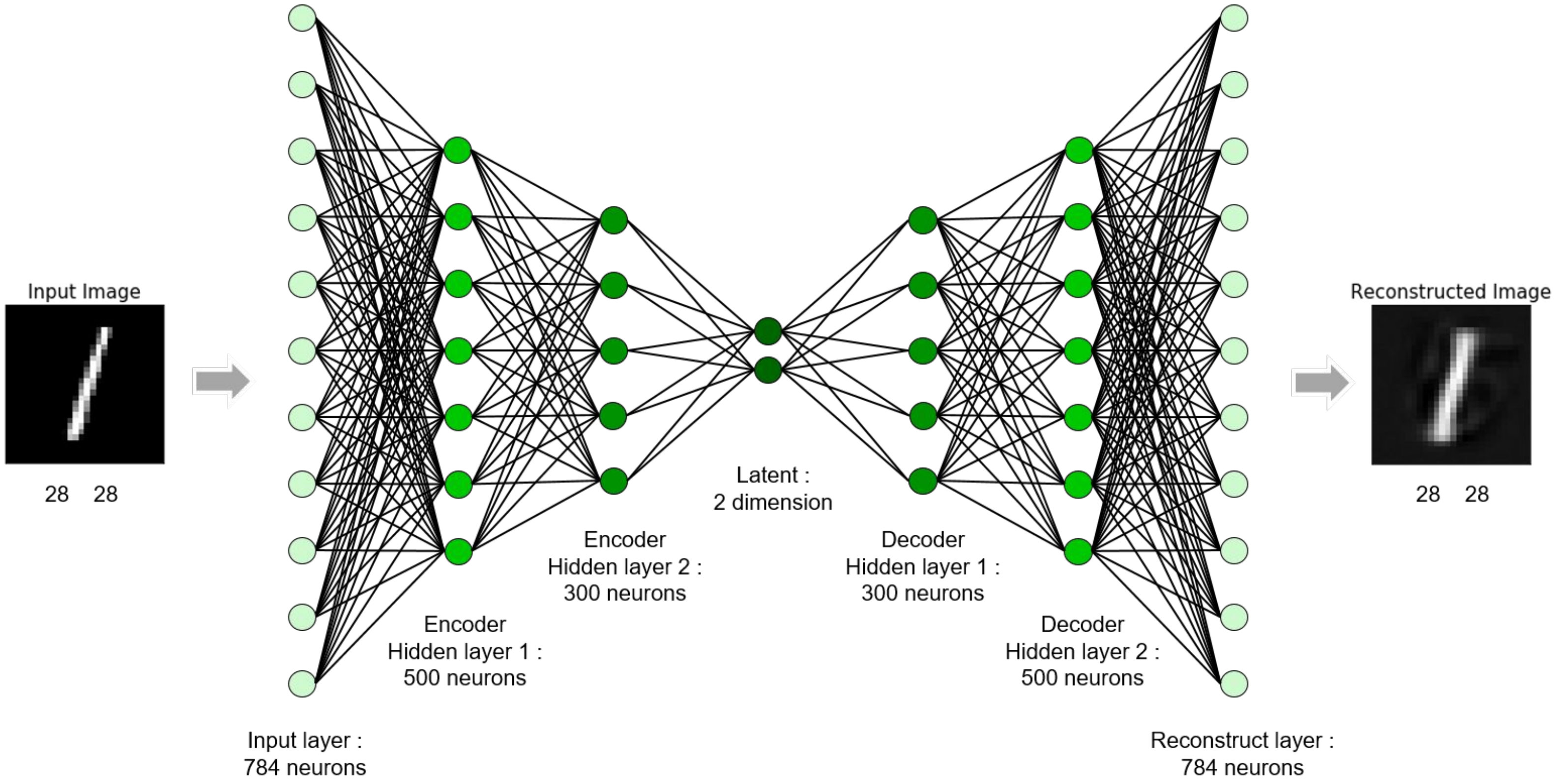


Malaria dataset
1 cell == 1 image
Task: infected vs. uninfected

*(Rajaraman et al., 2018)*

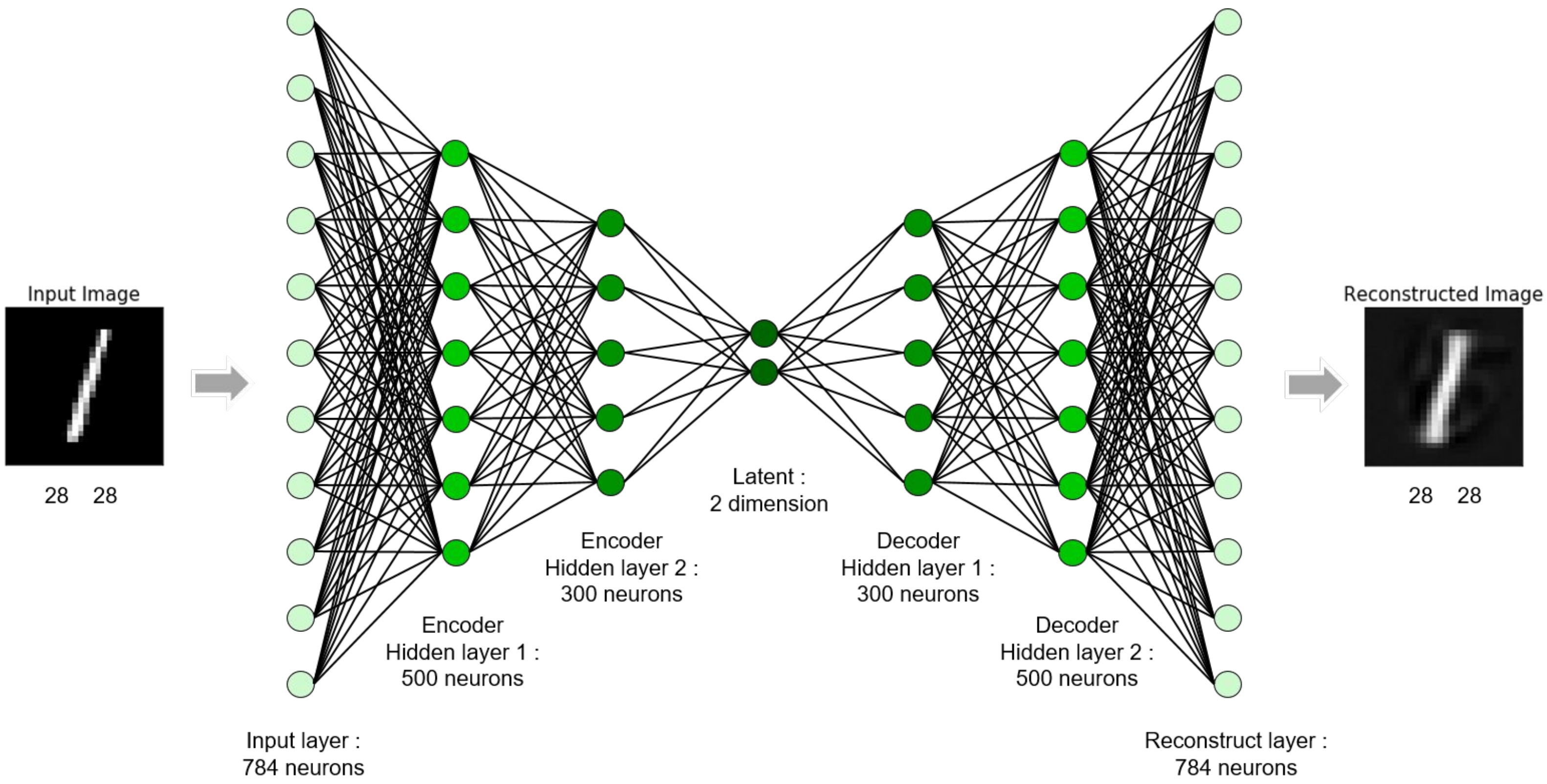
Can we disentangle the staining and the virus?

# Disentanglement




*(Kingma and Welling, 2014)*

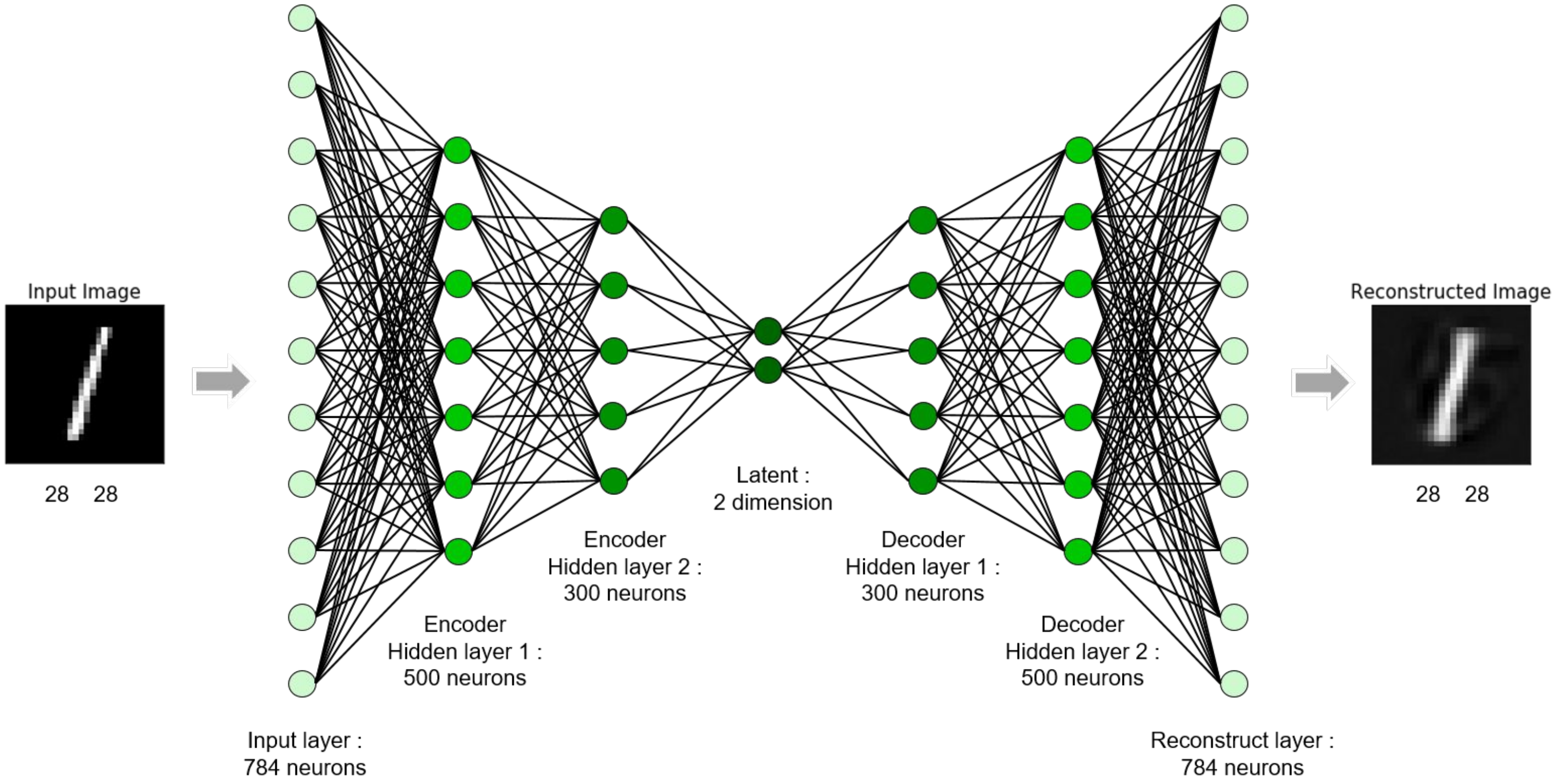
# Disentanglement




$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{x}^{(i)}) = \qquad + \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}^{(i)})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}|\mathbf{z})\right]$$

*(Kingma and Welling, 2014)*

# Disentanglement




$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{x}^{(i)}) = -D_{KL}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}^{(i)})||p_{\boldsymbol{\theta}}(\mathbf{z})) + \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}^{(i)})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}|\mathbf{z})\right]$$

*(Kingma and Welling, 2014)*

# Disentanglement


*(Kingma et al., 2014)*

# Disentanglement

Two latents:

- $z_1$ -> Content
- $z_2$ -> Style

(Kingma et al., 2014)

# Disentanglement

Two latents:

$z_1$ -> Content
$z_2$ -> Style

Changing one doesn't change the other

(Kingma et al., 2014)

# Disentanglement

Two latents:

$z_1$ -> Content
$z_2$ -> Style

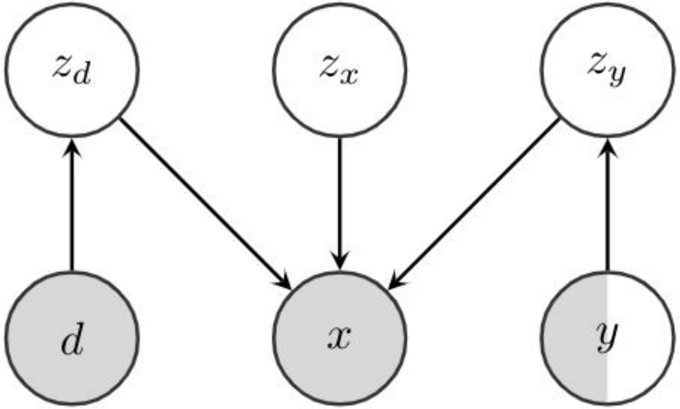
Changing one doesn't change the other

Idea: Just use $z_1$ for classification


(Kingma et al., 2014)

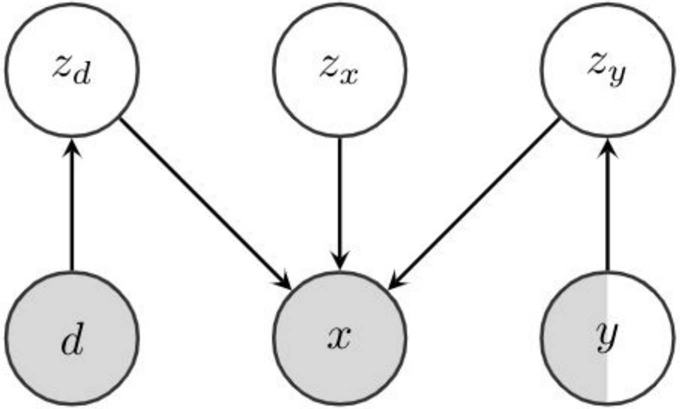
# DIVA

# DIVA




Generative

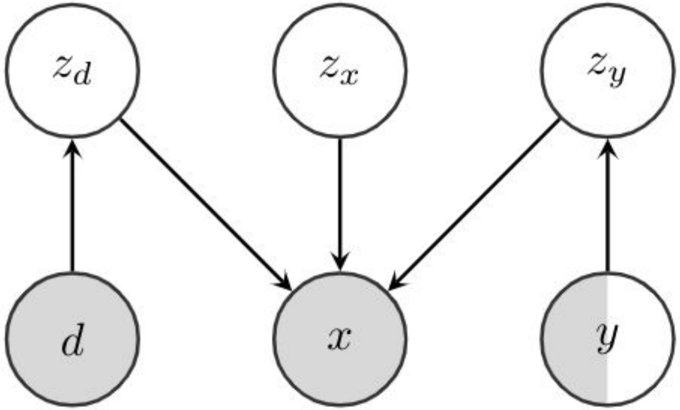
# DIVA



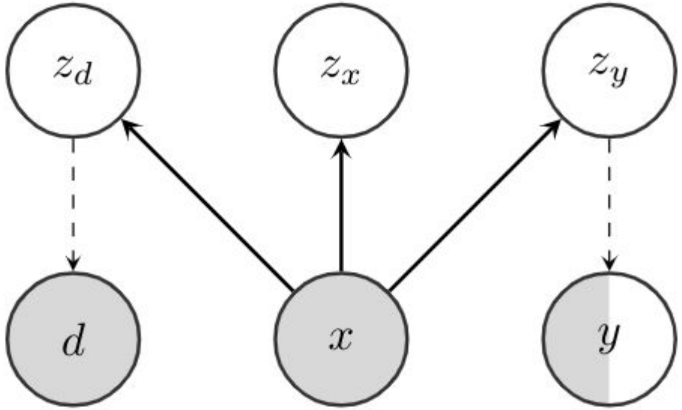

Generative

Think of: d = patient, x = cell, y = infected/uninfected -> training tuple (x, y, d)

# DIVA



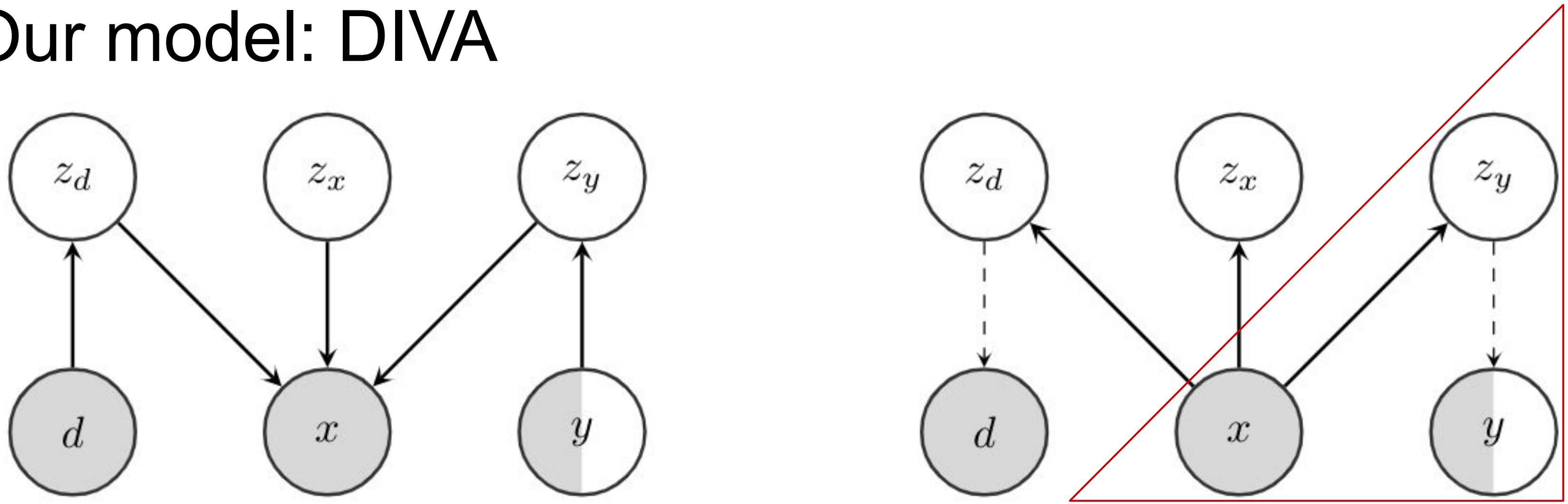

Generative




Inference

Think of: d = patient, x = cell, y = infected/uninfected -> training tuple (x, y, d)

# Our model: DIVA

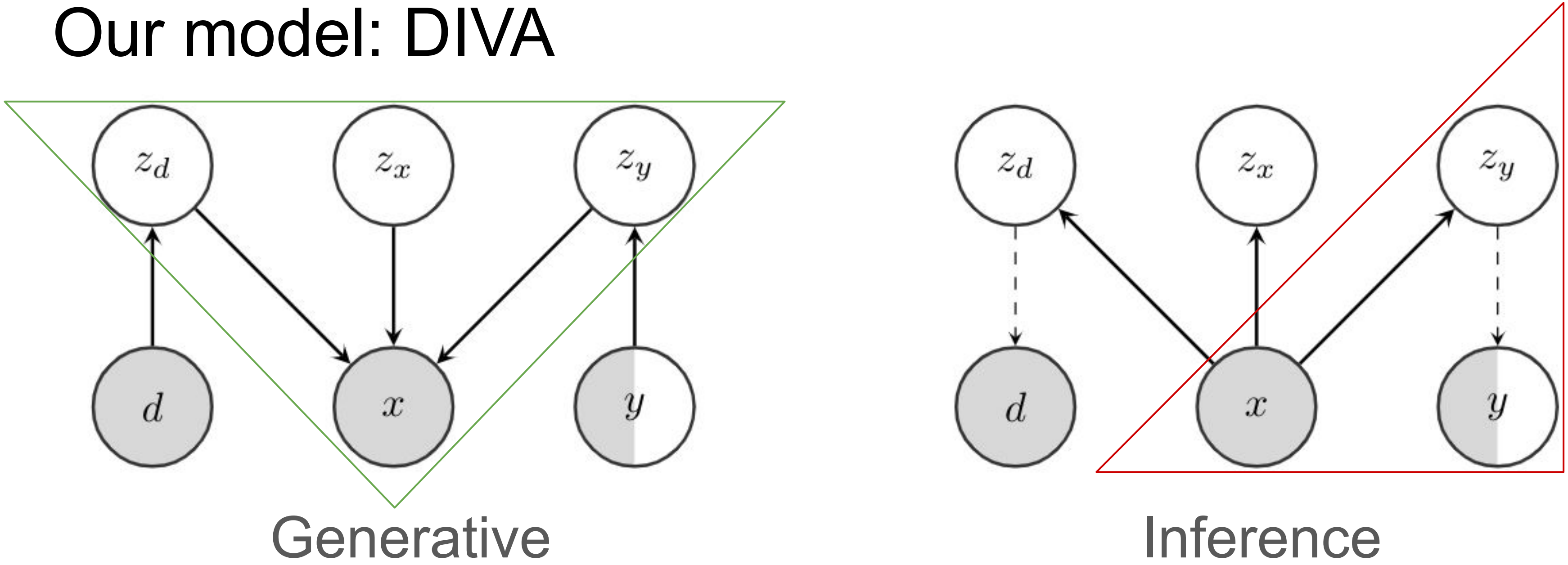



Generative Inference

Think of: d = patient, x = cell, y = infected/uninfected -> training tuple (x, y, d)

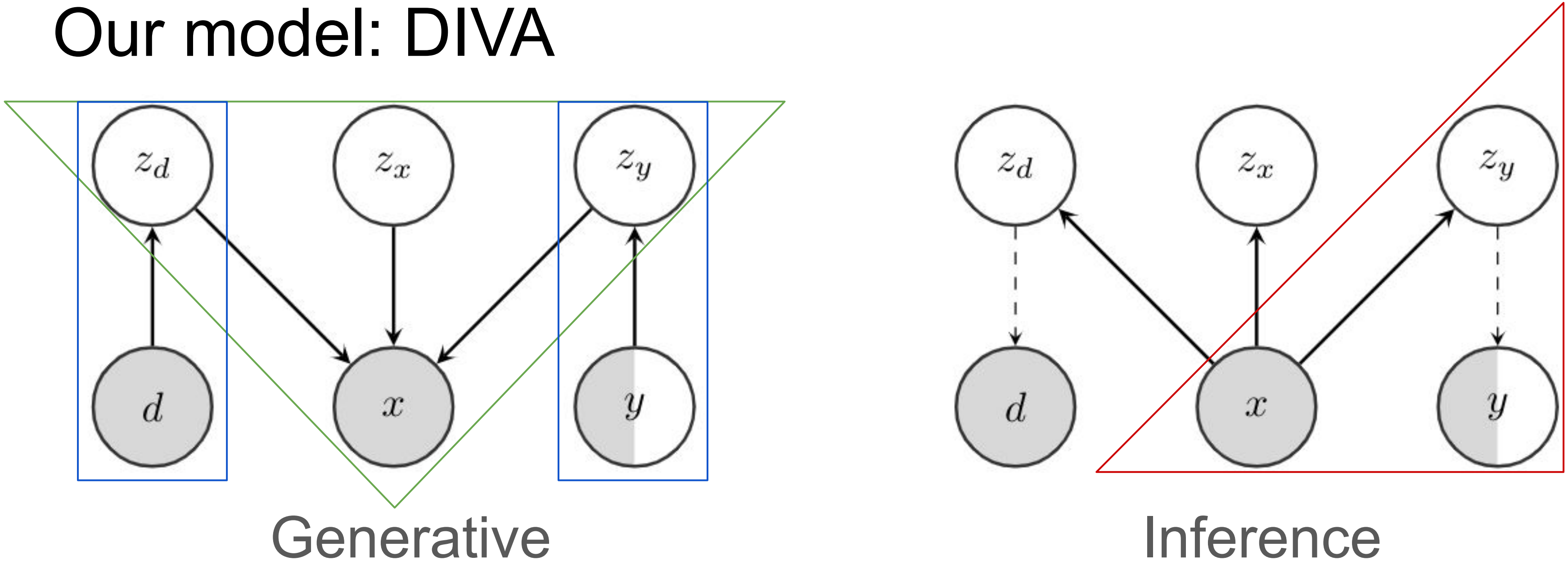
Red: CNN for classification of y, dashed arrows == auxiliary classifiers

# Our model: DIVA




Think of: d = patient, x = cell, y = infected/uninfected -> training tuple (x, y, d)
Red: CNN for classification of y, dashed arrows == auxiliary classifiers
Green: Reconstruction of x

# Our model: DIVA



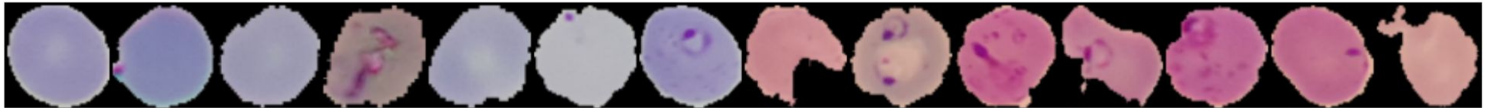

Think of: d = patient, x = cell, y = infected/uninfected -> training tuple (x, y, d)

Red: CNN for classification of y, dashed arrows == auxiliary classifiers

Green: Reconstruction of x

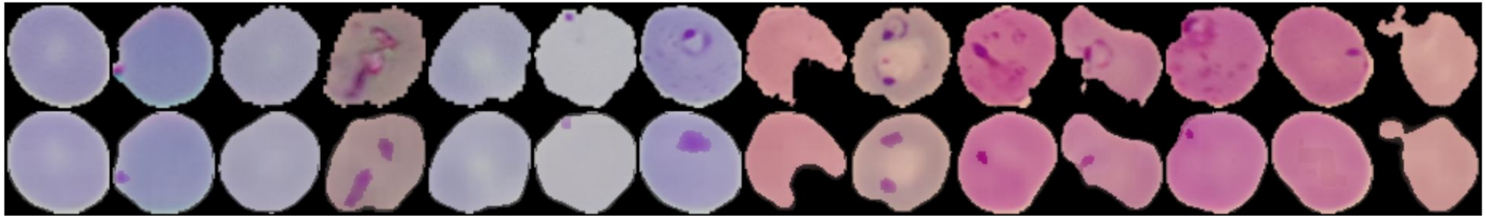
Blue: Conditional prior distributions

# Qualitative results



$\mathbf{x}$

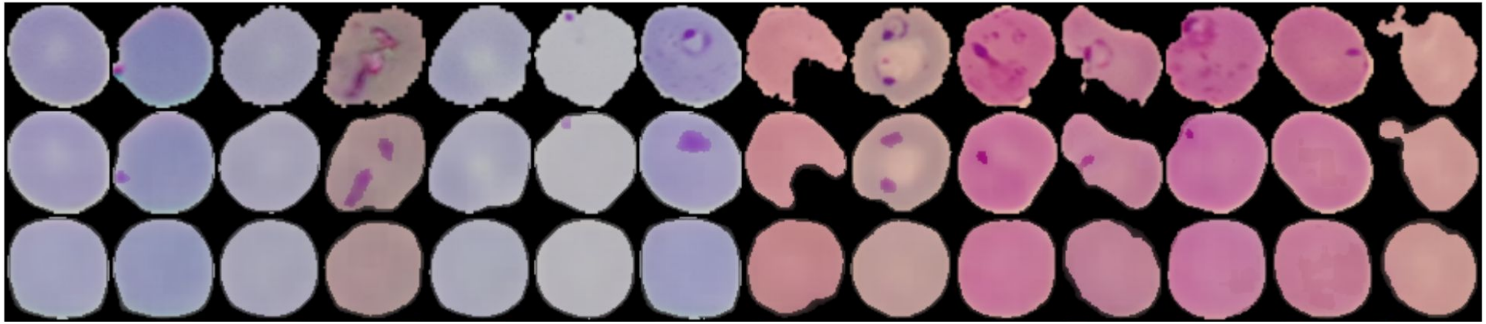
# Qualitative results



$\mathbf{x}$

$\mathbf{x} \sim p_\theta(\mathbf{x}|\mathbf{z}_d, \mathbf{z}_x, \mathbf{z}_y)$

# Qualitative results

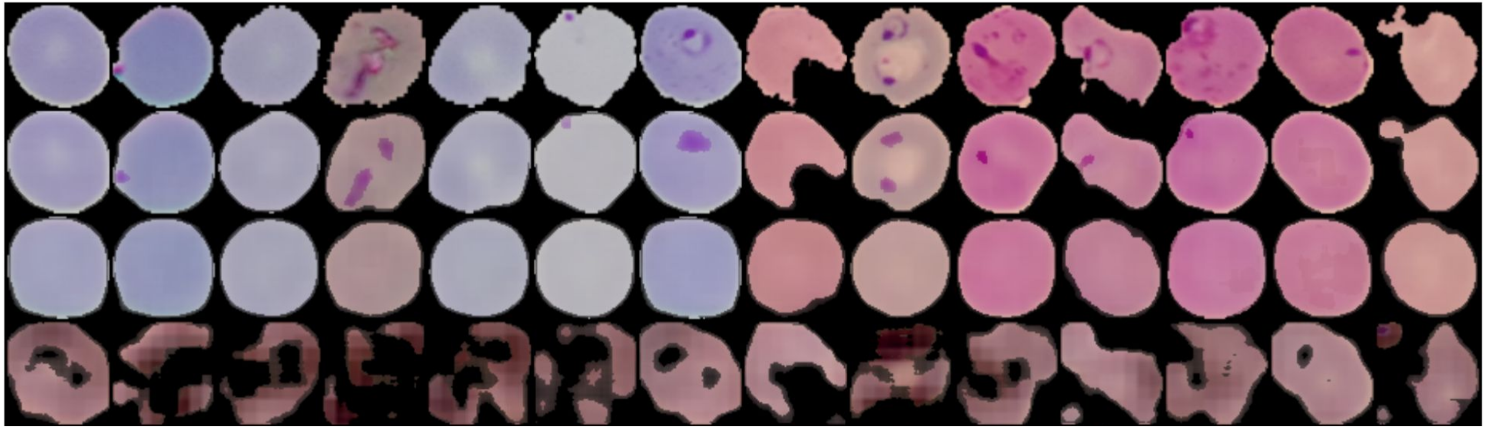


$\mathbf{x}$

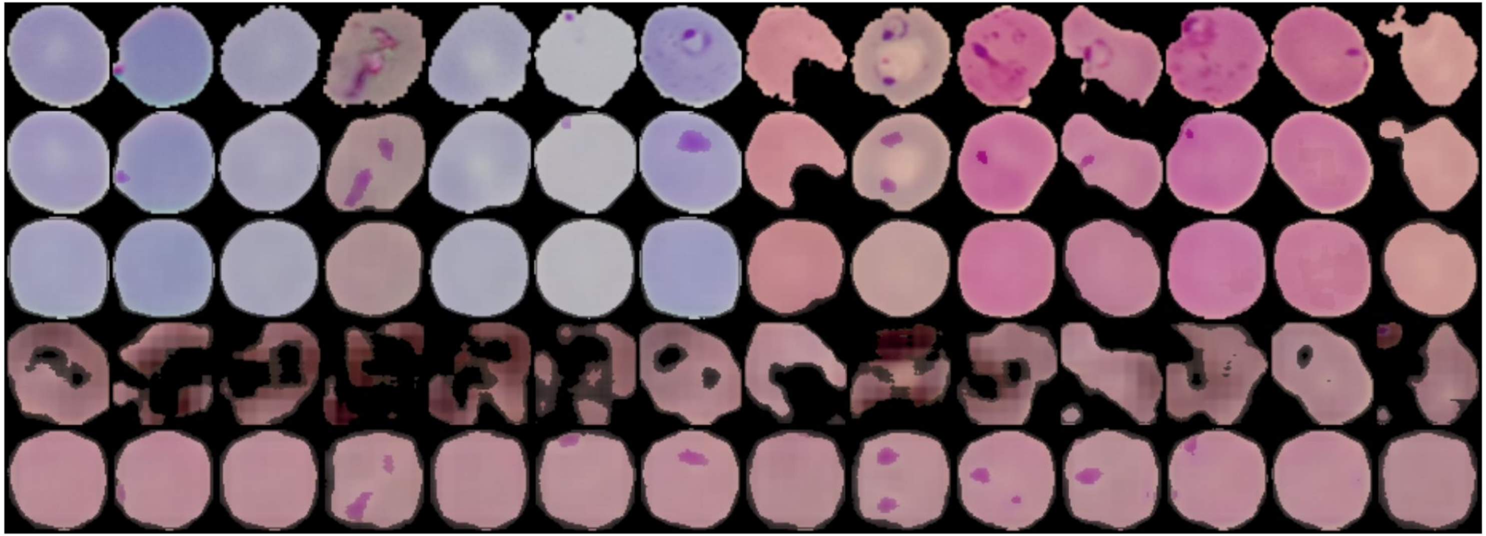
$\mathbf{x} \sim p_\theta(\mathbf{x}|\mathbf{z}_d, \mathbf{z}_x, \mathbf{z}_y)$

$\mathbf{x} \sim p_\theta(\mathbf{x}|\mathbf{z}_d, 0, 0)$

# Qualitative results



# Qualitative results



# Quantitative results

# Quantitative results

Table 3: Results of the supervised experiments for the first part of domains. We report the average and standard error of ROC AUC.

| Model | C116P77 | C132P93 | C137P98 | C180P141 | C182P143 | C184P145 |
|---|---|---|---|---|---|---|
| Baseline | $90.6 \pm 0.7$ | $97.8 \pm 0.5$ | $98.9 \pm 0.2$ | $98.5 \pm 0.2$ | $96.7 \pm 0.4$ | $98.1 \pm 0.2$ |
| DA | $90.6 \pm 1.7$ | $\mathbf{98.3 \pm 0.4}$ | $99.0 \pm 0.1$ | $98.8 \pm 0.1$ | $96.9 \pm 0.4$ | $97.1 \pm 0.8$ |
| DIVA | $\mathbf{93.3 \pm 0.4}$ | $\mathbf{98.4 \pm 0.3}$ | $99.0 \pm 0.1$ | $\mathbf{99.0 \pm 0.1}$ | $96.5 \pm 0.3$ | $\mathbf{98.5 \pm 0.3}$ |

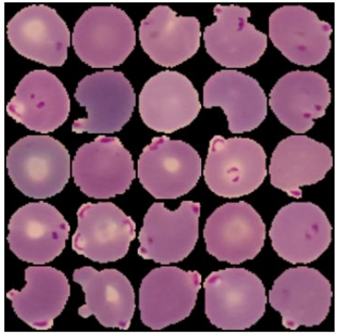
Table 4: Results of the supervised experiments for the second part of domains. As well as the average across all domains. We report the average and standard error of ROC AUC.

| Model | C39P4 | C59P20 | C68P29 | C99P60 | Average |
|---|---|---|---|---|---|
| Baseline | $97.1 \pm 0.4$ | $82.8 \pm 2.8$ | $95.3 \pm 0.6$ | $96.2 \pm 0.1$ | $95.2 \pm 1.6$ |
| DA | $97.4 \pm 0.3$ | $83.2 \pm 3.3$ | $\mathbf{96.3 \pm 0.1}$ | $96.1 \pm 0.3$ | $95.4 \pm 1.6$ |
| DIVA | $\mathbf{97.8 \pm 0.2}$ | $82.1 \pm 3.0$ | $\mathbf{96.3 \pm 0.2}$ | $\mathbf{96.6 \pm 0.3}$ | $95.8 \pm 1.6$ |

# Unsupervised domains

# Unsupervised domains

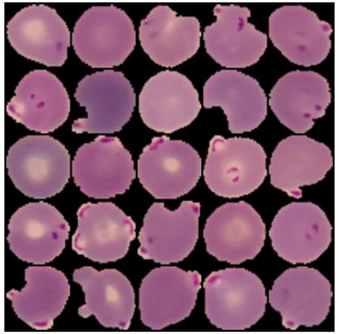
If I want to generalise to this patient



(a) C116P77

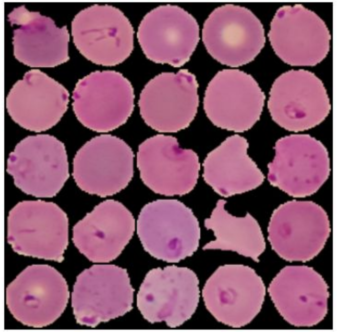
# Unsupervised domains

If I want to generalise to this patient



(a) C116P77

Does it help to have unlabeled data from this patient

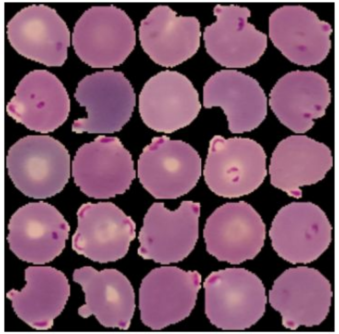


?

(h) C59P20

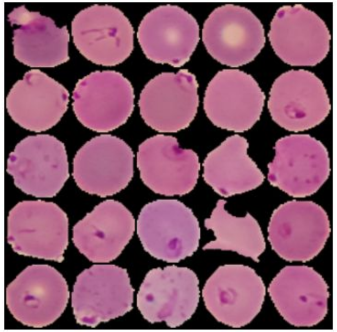
# Unsupervised domains

If I want to generalise to this patient



(a) C116P77

Does it help to have unlabeled data from this patient



?

(h) C59P20

| Training data | Baseline | DA | DIVA |
|---|---|---|---|
| Labeled data from C59P20 | $90.6 \pm 0.7$ | $90.6 \pm 1.7$ | $\mathbf{93.3 \pm 0.4}$ |
| Unlabeled data from C59P20 | - | $72.05 \pm 2.2$ | $\mathbf{79.4 \pm 2.8}$ |
| No data from C59P20 | $70.0 \pm 2.6$ | $69.2 \pm 1.9$ | $71.9 \pm 2.7$ |

Thank you for your attention!