Compressive Visual Representations

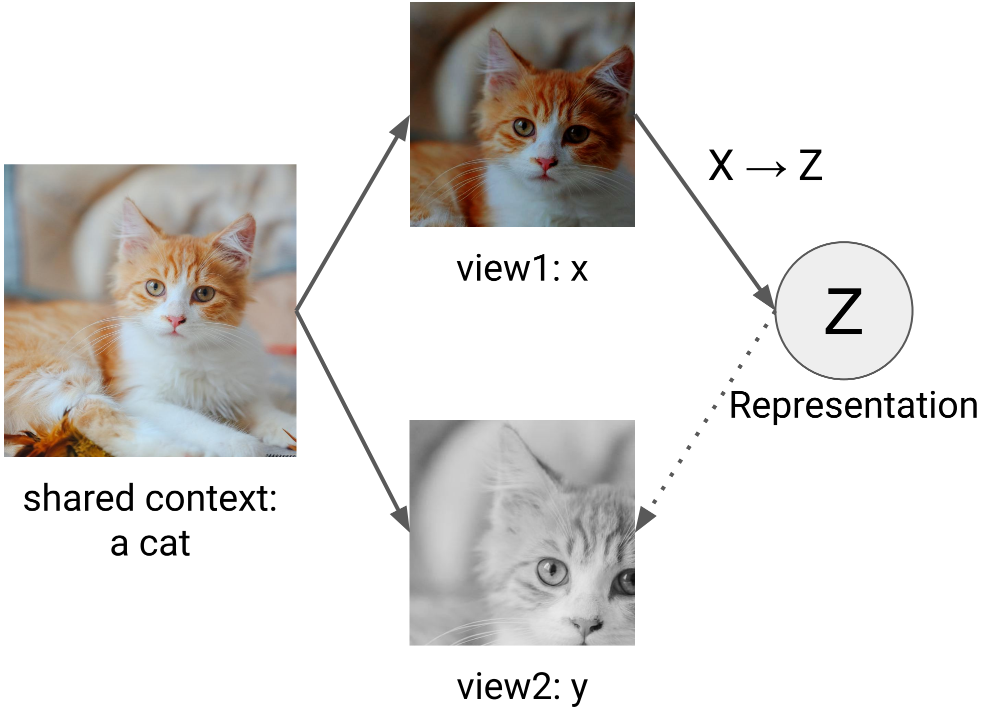
Kuang-Huei Lee†, Anurag Arnab†,
Sergio Guadarrama, John Canny, Ian Fischer†

†: Main contributors

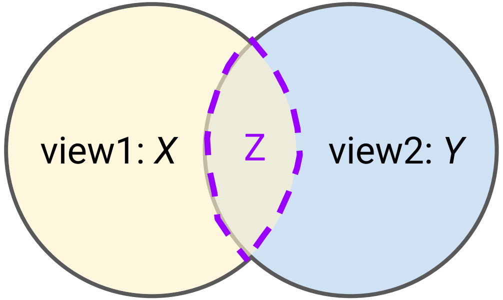
Google Research

The multiview self-supervised learning framework

Capture the shared invariant between different views

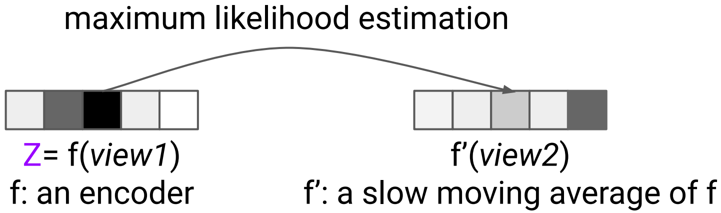


InfoMax approaches (e.g. contrastive methods)



Desired outcome: I(X;Z) = I(Y;Z) = I(X;Y)

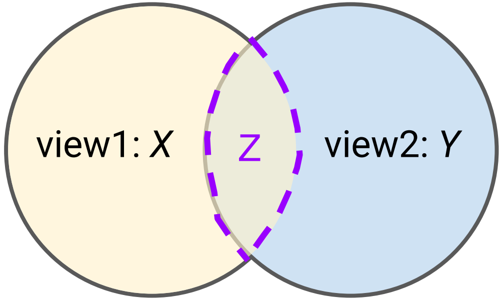
Bootstrap Your Own Latent (BYOL) approaches



InfoMax and maximum likelihood are NOT ideal

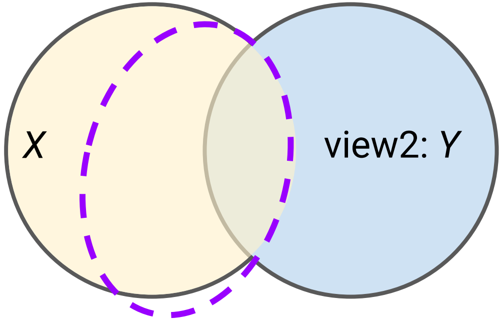
Most contrastive and latent bootstrapping methods did not consider this

What all multiview self-supervised methods hope to achieve:



A representation Z that captures only the invariant, i.e. I(X;Y)

What InfoMax and maximum likelihood actually achieve by maximizing I(Y;Z):



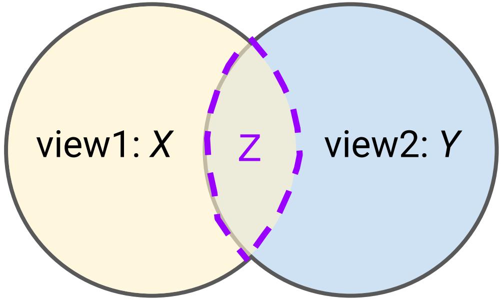
Any Z that contains I(X;Y) is valid at optimal, no matter it contains how much irrelevant info about X

This Z is obviously not the invariant you want.

InfoMax and maximum likelihood are NOT ideal

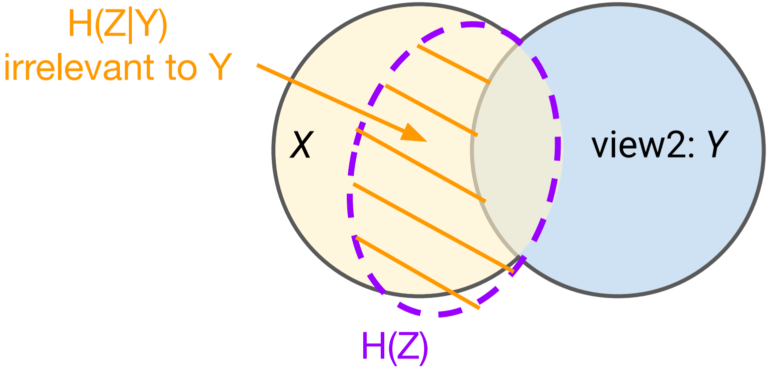
Most contrastive and latent bootstrapping methods did not consider this

What all multiview self-supervised methods hope to achieve:



A representation Z that captures only the invariant, i.e. I(X;Y)

What InfoMax and maximum likelihood actually achieve by maximizing I(Y;Z):



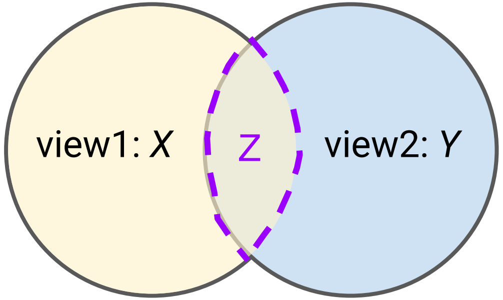
Any Z that contains I(X;Y) is valid at optimal, no matter it contains how much irrelevant info about X

This Z is obviously not the invariant you want.

InfoMax and maximum likelihood are NOT ideal

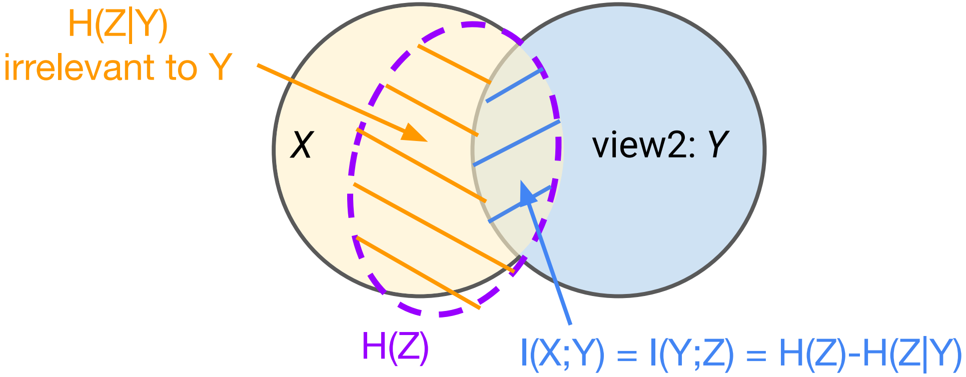
Most contrastive and latent bootstrapping methods did not consider this

What all multiview self-supervised methods hope to achieve:



A representation Z that captures only the invariant, i.e. I(X;Y)

What InfoMax and maximum likelihood actually achieve by maximizing I(Y;Z):



Any Z that contains I(X;Y) is valid at optimal, no matter it contains how much irrelevant info about X

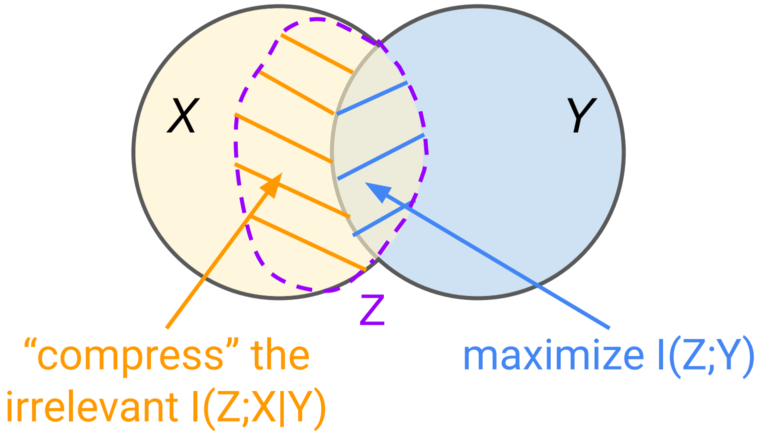
This Z is obviously not the invariant you want.

The Conditional Entropy Bottleneck (CEB)

Explicitly compress self-supervised models with the CEB objective

Compressing a self-supervised model is very simple!

1. Convert the final output representations into explicit distributional forms
2. Replace the InfoMax or maximum likelihood objective with CEB



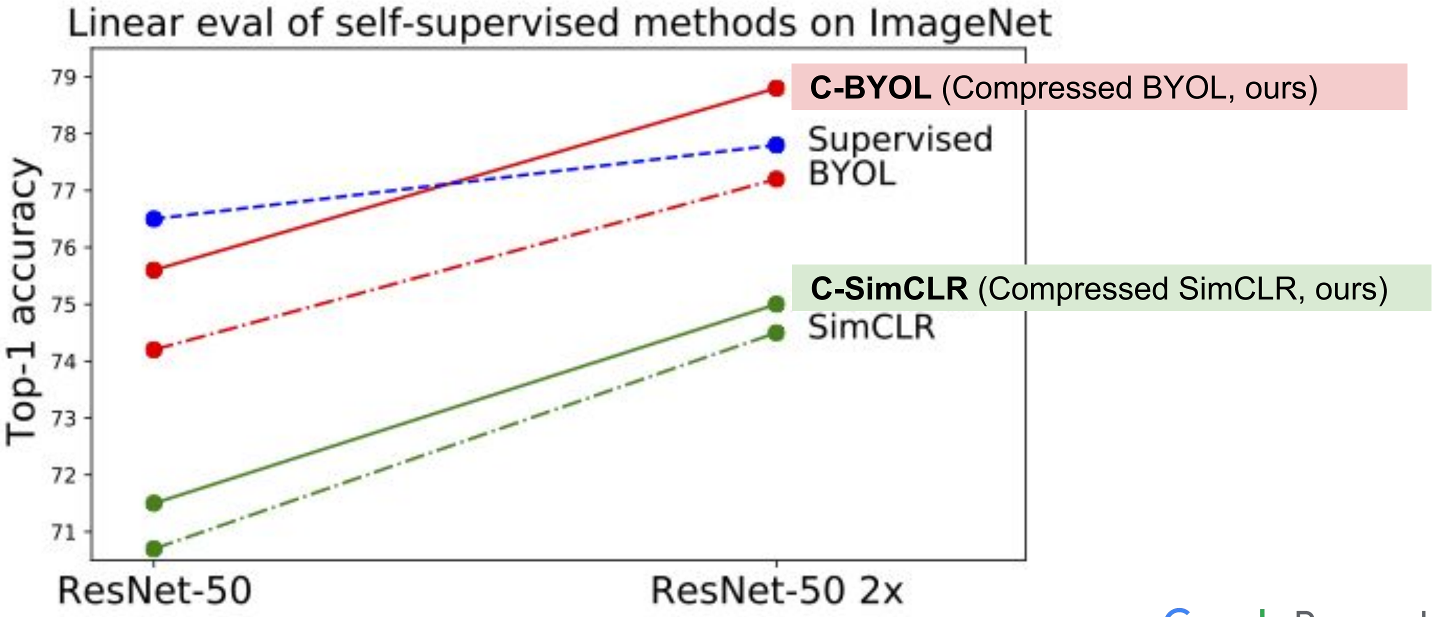
$$
\begin{aligned}
CEB &\equiv \min_{Z} \beta I(X;Z|Y) - I(Y;Z) \\
&\equiv \min_{Z} \beta(-H(Z|X) + H(Z|Y)) + H(Y|Z)
\end{aligned}
$$

There is a tractable variational bound for CEB:

$$
vCEB \equiv \min_{e(z|x),b(z|y),d(y|z)} \mathbb{E}_{x,y\sim p(x,y),z\sim e(z|x)} \beta(\log e(z|x) - \log b(z|y)) - \log d(y|z)
$$

Milestone: supervised-level, self-supervised results

We train a linear classifier on the self-supervised representation, learned without labels.
It achieves results as good as fully supervised models, showing how general the representation is.



Google Research

Comparison is system-wise as in prior work. We use reasonable supervised baselines reported in the SimCLR and BYOL papers.

Compressed models are robust to out-of-distribution

github.com/google-research/robustness_metrics

Method	ImageNet-A	ImageNet-C	ImageNet-R	ImageNet-v2	ImageNet-Vid	YouTube-BB	ObjectNet
SimCLR	1.3	35.0	18.3	57.7	63.8	57.3	18.7
C-SimCLR	**1.4**	**36.8**	**19.6**	**58.7**	**64.7**	**59.5**	**20.8**
BYOL	1.6	42.7	24.4	62.1	67.9	60.7	23.4
C-BYOL	**2.3**	**45.1**	**25.8**	**63.9**	**70.8**	**63.6**	**25.5**

ImageNet-C
Common Corruptions

ObjectNet
Changing viewpoints & backgrounds

*The main paper introduces a new theoretical connection between CEB compression and the model's Lipschitz constant, helping to explain why compressed models are more robust.

Google Research

Compressed feature representations transfer better

Transfer self-supervised representations pre-trained on ImageNet to other classification tasks

Method	Food101	CIFAR10	CIFAR100	Flowers	Pet	Cars	Caltech-101	DTD	SUN397	Aircraft	Birdsnap
SimCLR	72.5	91.1	74.4	88.4	83.5	49.7	89.5	72.5	61.8	51.6	35.4
C-SimCLR	**73.0**	**91.6**	**75.2**	**89.0**	**84.0**	**52.7**	**91.2**	**73.0**	**62.3**	**53.5**	**38.2**

Thank you!

arxiv.org/abs/2109.12909

Our implementation is available on github

For questions, reach out to Kuang-Huei Lee (leekh@google.com)

Google Research