# Bootstrap Your Own Variance

Polina Turishcheva*[,1], Jason Ramapuram*[,2], Sinead Williamson*[,2], Dan Busbridge[2], Eeshan Dhekane[2], Russ Webb[2]
University of Göttingen[1], Apple Inc[2].

## Abstract

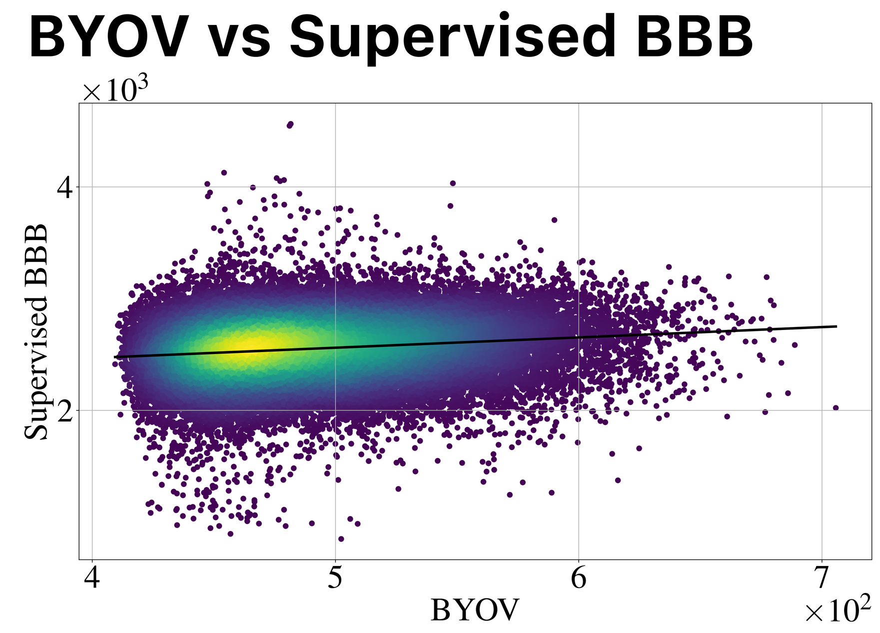
### Uncertainty Estimation

- Estimation of uncertainty is important for many applications.
- Uncertainty estimation typically requires labels.
- Can we estimate epistemic uncertainty without labels such that there exists a distributional relationship with the model trained with labels?

### Background

- Bayes-by-backprop (BBB) assumes a parametric tractable distribution $q(w|\theta)$ and maximizes an evidence lower bound (ELBO).
- BYOL is a distillation based self-supervised learning approach that learns an augmentation invariant latent representation.

### BYOV vs Supervised BBB

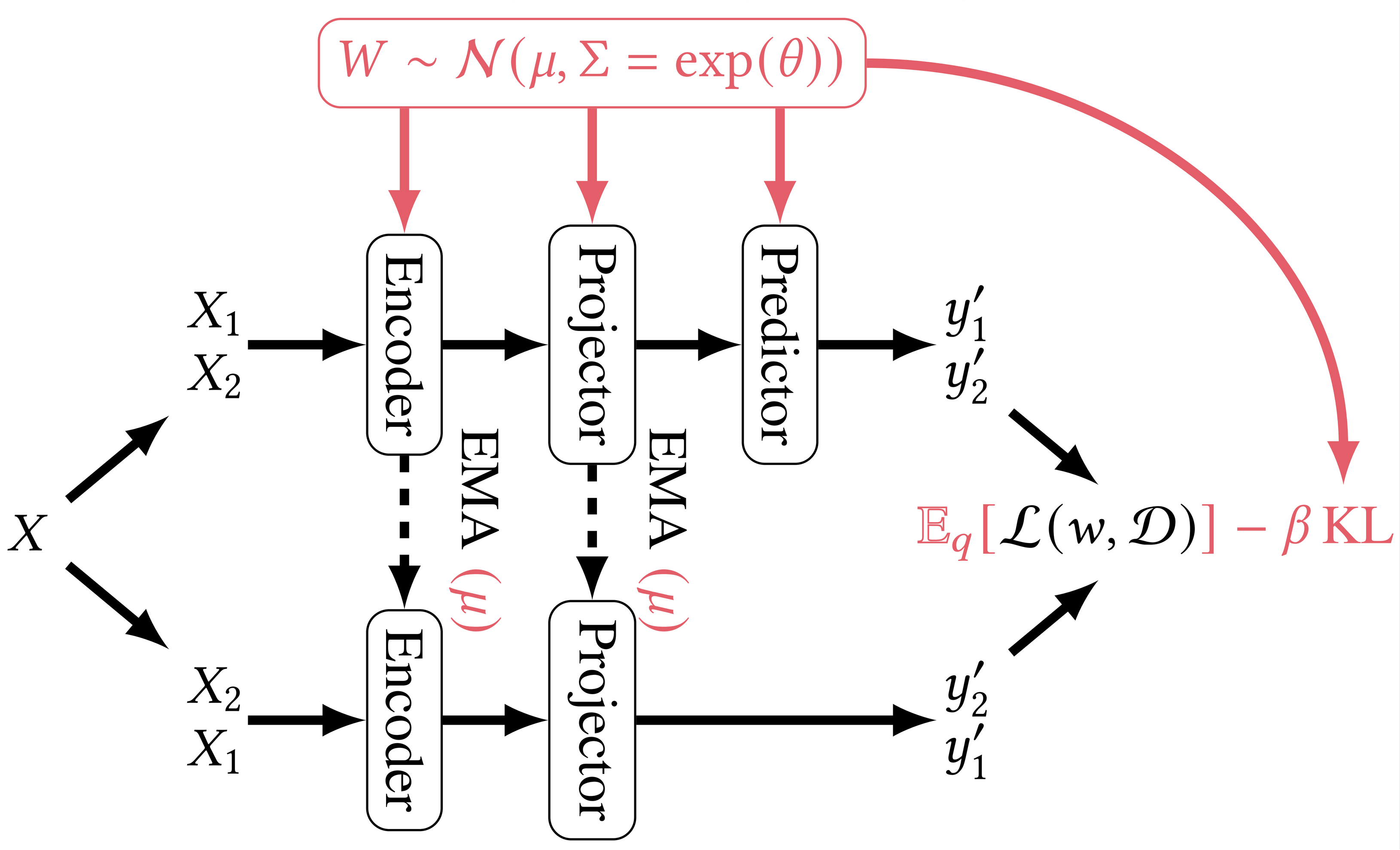



Variance between BYOV & Supervised BBB can be modeled as a Gaussian.

## Methods

### BYOV Architecture

BYOL + BBB with an approximate parameter posterior



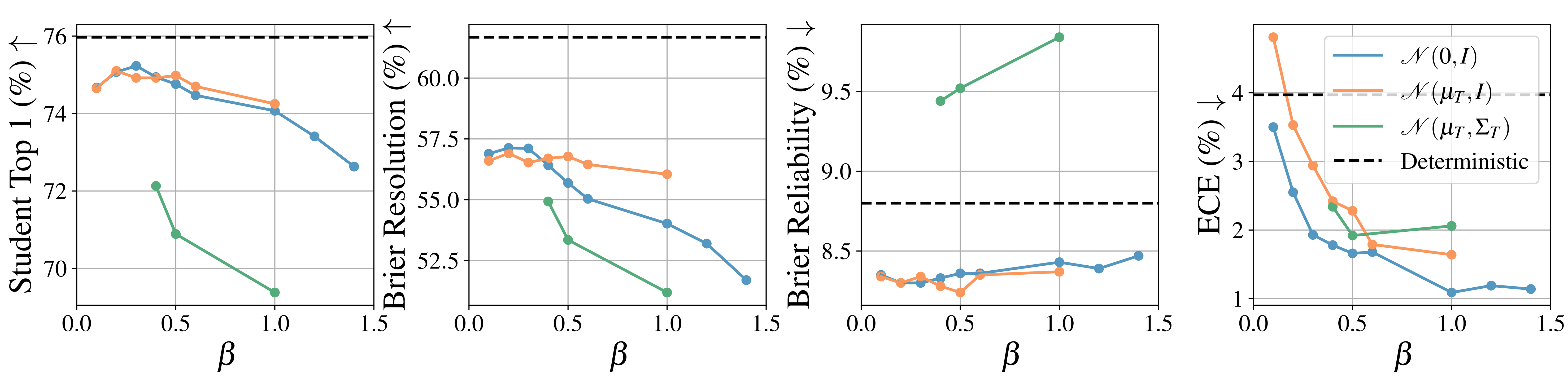

### Generalized Evidence Lower Bound

SSL with an ELBO

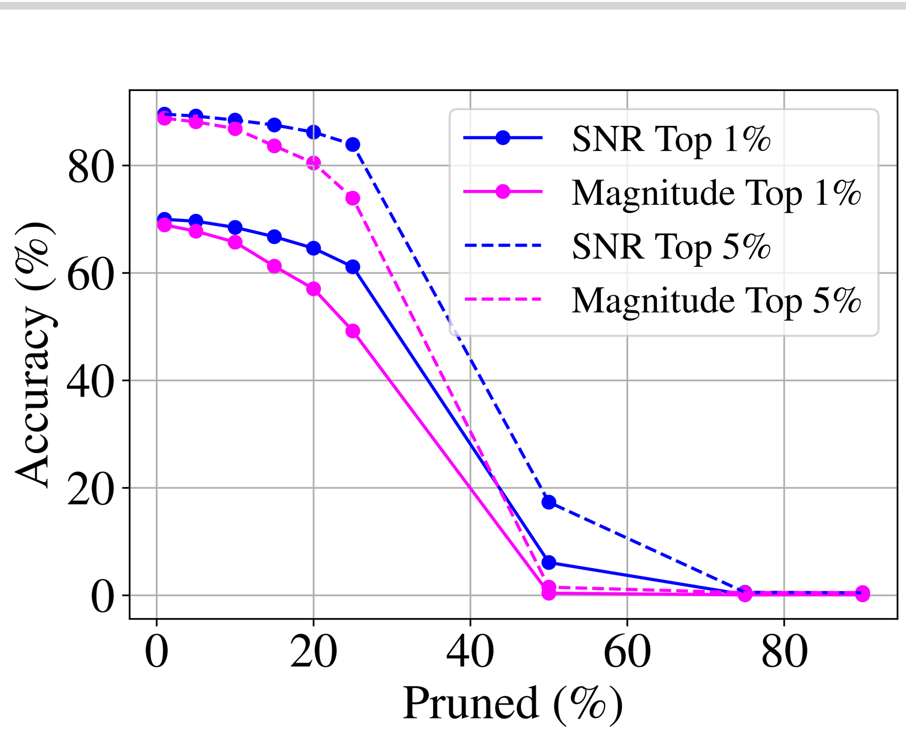
$$\mathbb{E}_{q(w|\theta)}[\underbrace{\mathcal{L}(w, \mathcal{D})}_{\text{cosine similarity}}] - \beta\ \mathrm{KL}[q(w|\theta)\,||\,p(w)]$$

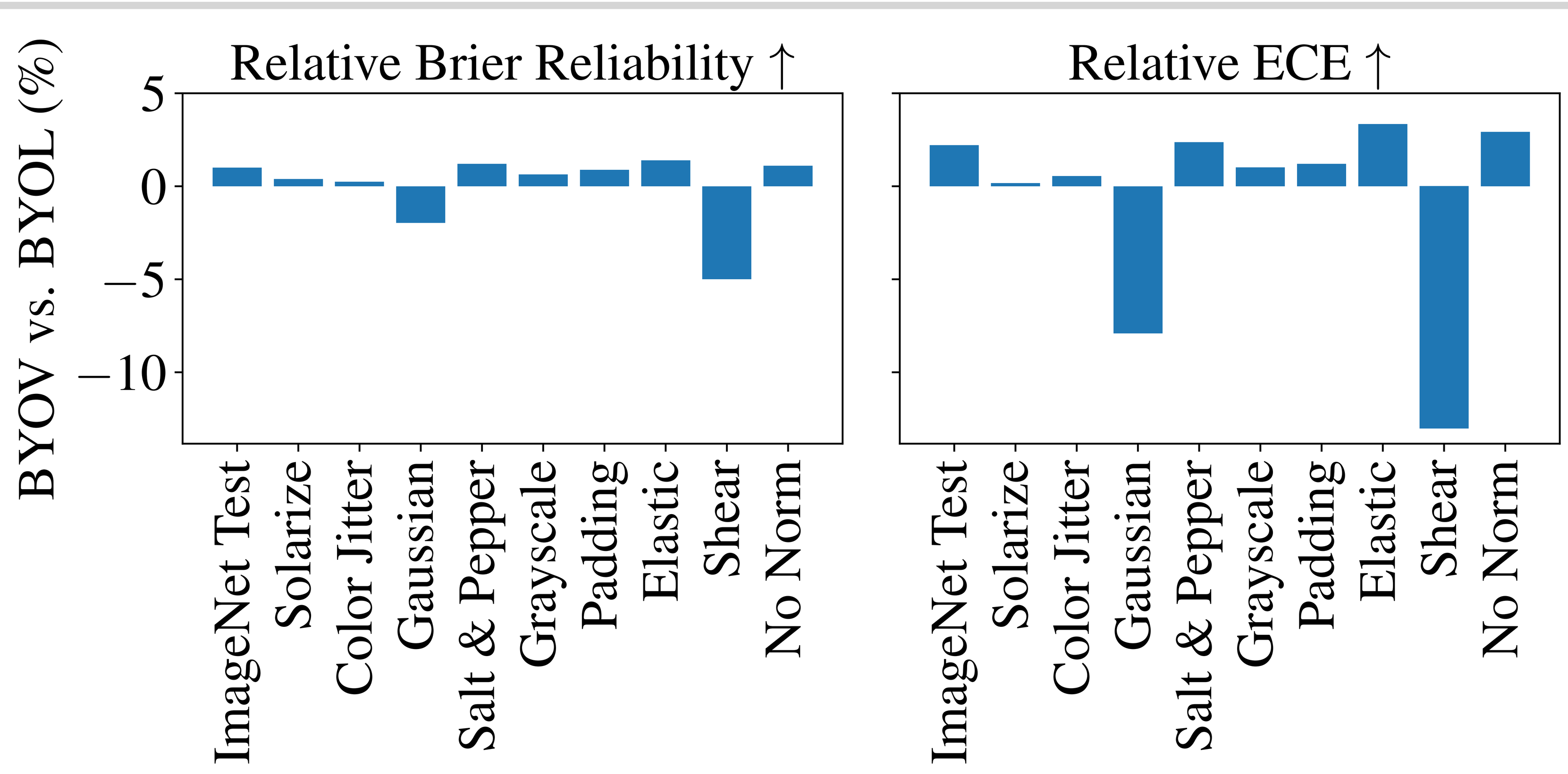
- BYOL minimizes a cosine similarity against an exponential moving average teacher.
- Cosine similarity does not prescribe a likelihood.
- BYOV bootstraps theory from generalized posterior and use a generalized likelihood.

## Results




- BYOV outperforms deterministic BYOL for ECE (**+2.83%**) and reliability (**+1.03%**).
- BYOV underperforms linear probing (-0.4%) and resolution (-0.57%) against deterministic BYOL.




- Left: BYOV enables signal-to-noise ratio pruning which outperforms magnitude based pruning (without retraining).
- Right: For OOD augmentations BYOV outperforms BYOL everywhere except shearing and additive Gaussian noise.