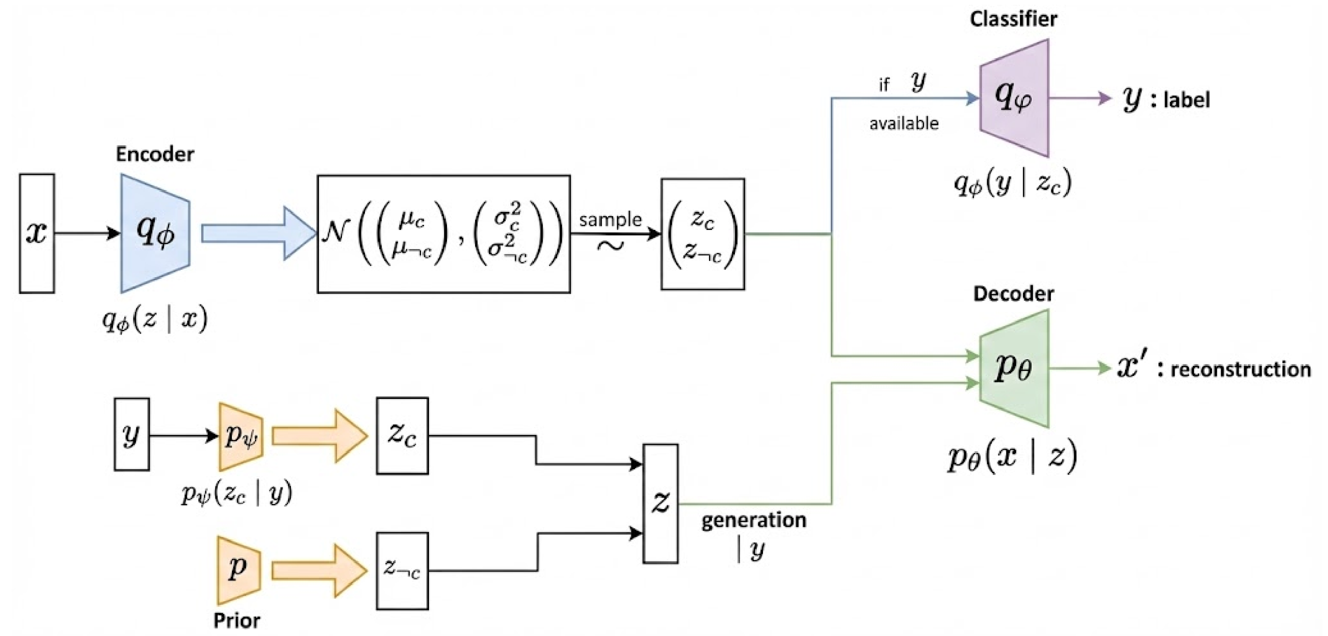
DO CCVAEs Practice What They Preach?

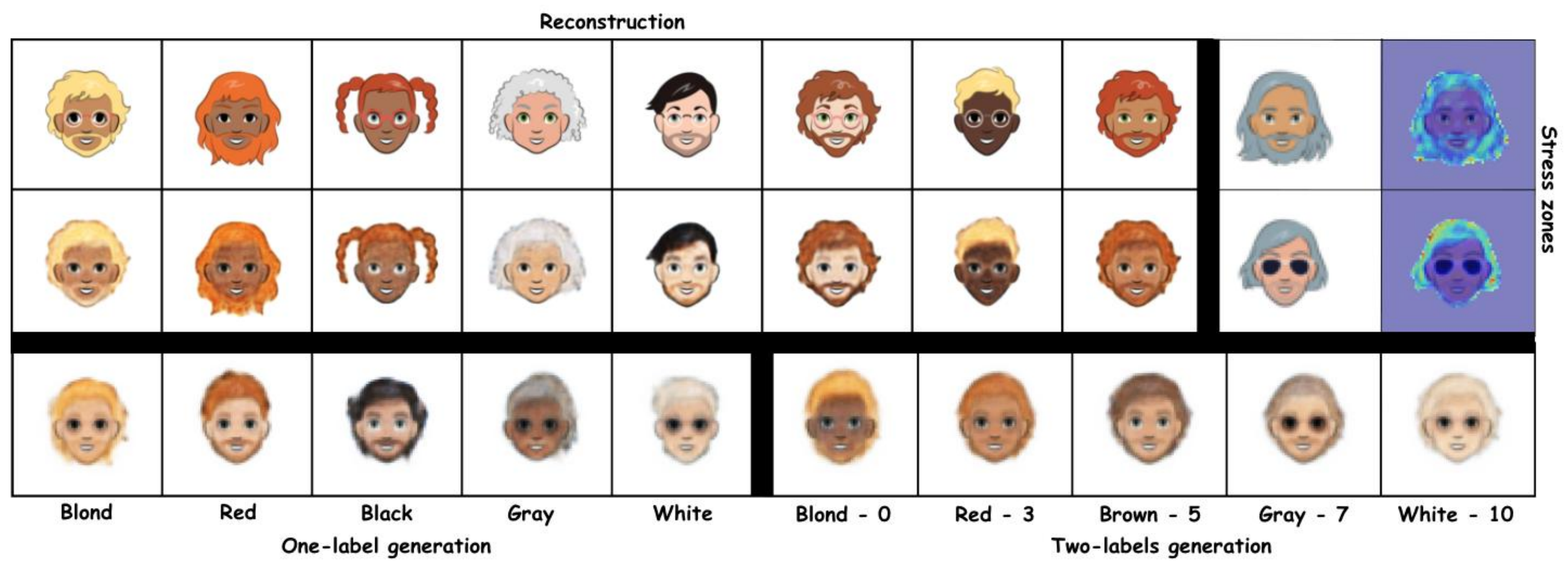
El Mehdi Nezahi, Farouk Yartaoui, Rida Assalouh

From input to generation in a CCVAE



- A Semi-Supervised approach extending standard VAEs.
- **Goal**: Split the latent space into z_c (Label-dependent content) and $z_{\neg c}$ (Residual style features).
- **Capabilities**: Classification, Regression, and Conditional Generation.

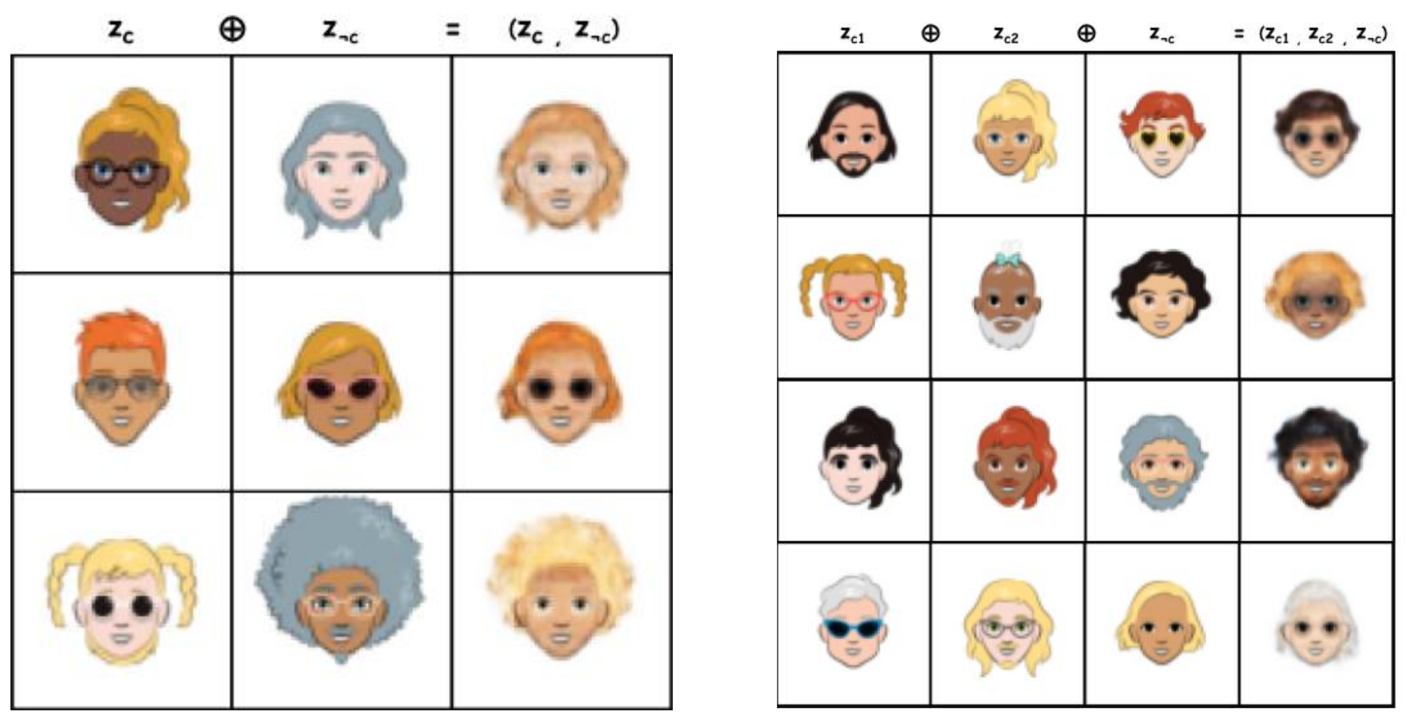
CCVAEs in Action: Reconstruction and Generation



Frankenstein test: Who Controls What?

•Mixing z_c components from different images while keeping a shared $z_{\neg c}$

$$z_{c1} \oplus z_{c2} \oplus \cdots \oplus z_{cN} \oplus z_{\neg c}$$



From Discrete Labels to Continuous Targets (C3VAE)

Benchmarking performance

- Comparing CCVAEs against classical baselines on classification and regression tasks.
- CCVAEs achieve competitive accuracy while retaining generative capabilities.

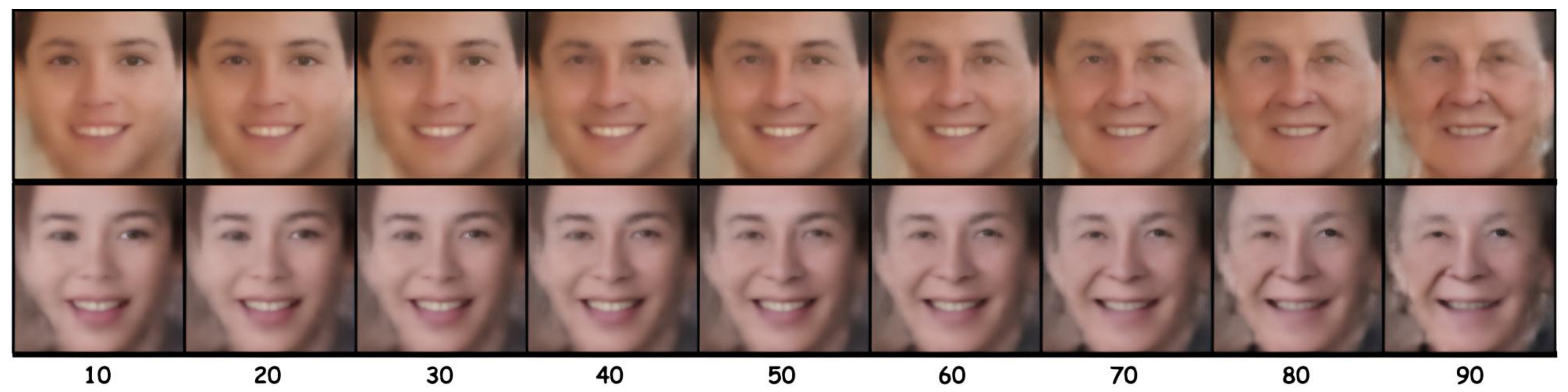
Task			CCVAE	CNN	SVM	LoReg*	RanFor*	KNN
Classif cation	1-label		95.1%	96.2%	95.9%	98%	98.5%	77%
	2-label	hair	93.93%	92.05%	96.54%	97.58%	85.20%	68.76%
		face	80.99%	65.70%	99.98%	98.84%	45.66%	35.40%
Regression			7.31	8.49	17.79	11.19	13.20	13.53

* LoReg: Logistic Regression
* RanFor: Random Forest

Continuous control with C3VAE

- Extending CCVAEs to regression task : Age prediction on real faces.
- Maintaining reconstruction and generation abilities.

- **Latent Traversal:** Fixing $z_{\neg c}$ and using conditionally generated z_c from 10 to 90 years old.
- **Result:** Smooth aging progression while retaining identity.

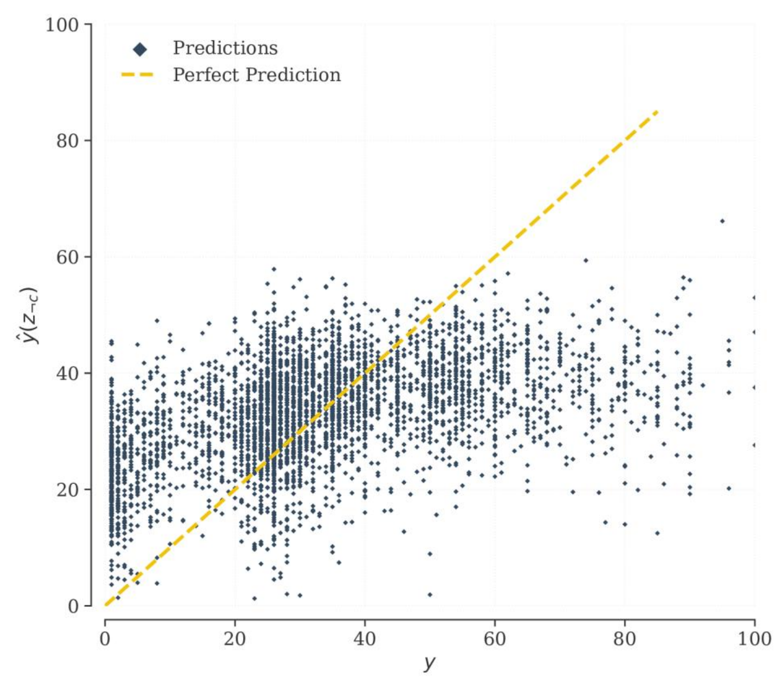


Latent Space Analysis

Information Leakage ?

- Probing age information against recovered $z_{\neg c}$, from the previous regression task.
- Age can be partially predicted from $z_{\neg c}$ ($R^2 = 0.30$).

⇒ **Unintended information leakage.**



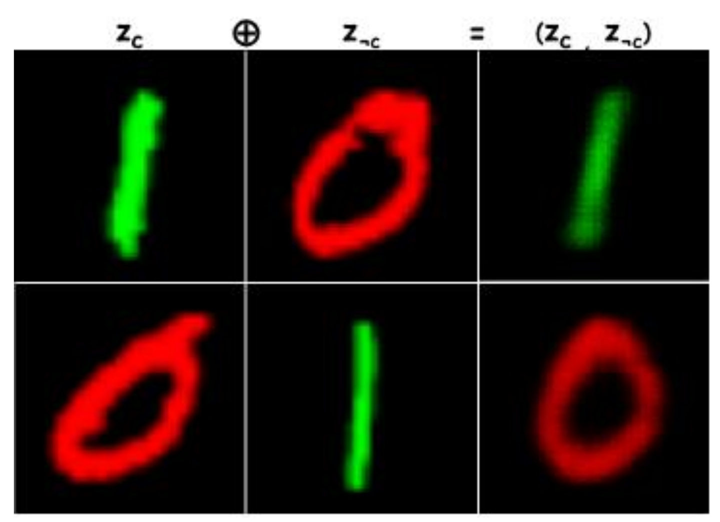
Irrelevant information ?

- Glasses prediction from z_c and $z_{\neg c}$ on CartoonSet (CCVAE trained on hair and face colors).
- z_c contains more information about glasses than $z_{\neg c}$.

⇒ **Lack of disentanglement.**

Classif er	Accuracy (%)	
	$z_{\neg c}$	z_c
Logistic Regression	62.4%	62.9%
XGBoost	65.9%	67.3%

Leaked correlations ?



- Forcing 90% correlation: Red zeroes and Green ones.
- Frankenstein test
  - **Desired:** Red One and Green Zero
  - **Obtained:** Green One and Red Zero

⇒ **Double-Edged Sword**

Discovering patterns in $z_{\neg c}$

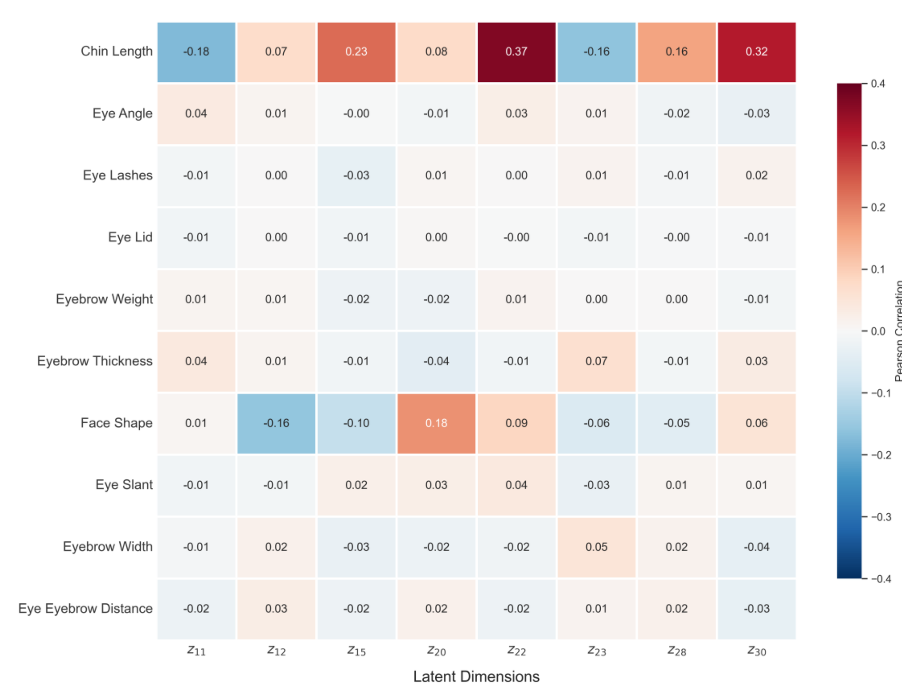
Emerging features from latent dimensions

- Looking for dimension-wise correlations with unseen labels.

$$\mathrm{Corr}(z_{\neg c}[k], y_\ell) = \frac{\mathrm{Cov}(z_{\neg c}[k], y_\ell)}{\sqrt{\mathrm{Var}(z_{\neg c}[k])\,\mathrm{Var}(y_\ell)}}$$

- Strong correlations emerge: concentrated information in single dimensions.

⇒ **z_{22} is related to chin length. Does it control it?**

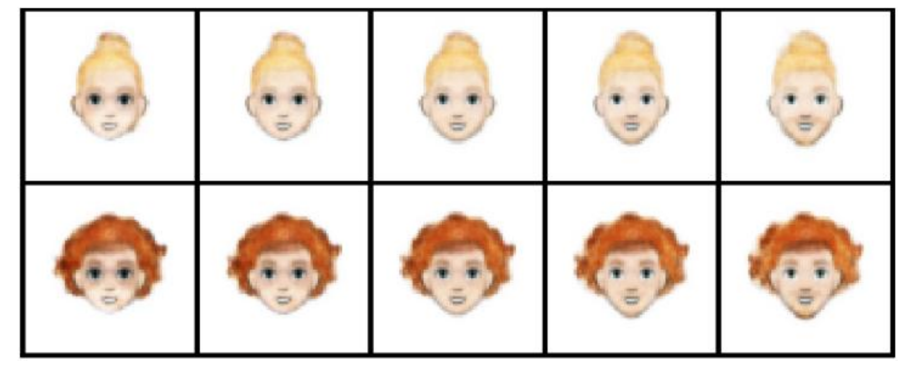


- For $\alpha \in [-10{,}10]$:

$$z_{\neg c} \leftarrow z_{\neg c} + \alpha\, e_{22}, \qquad e_{22} = (0,\ldots,0,1,0,\ldots,0)$$

- Decoding $(z_c, z_{\neg c})$: the variable α strongly impacts the chin length.

⇒ **Yes, it does! Small correlation but high impact.**



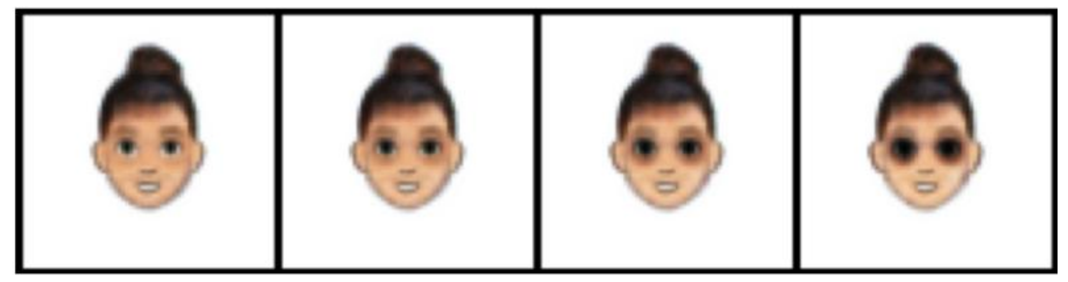
Emerging features from latent directions

- Glasses latent representations can't be explained by a single latent dimension.
- What about an aggregate direction ? → **Logistic regression could tell.**

$$\hat{\hat{y}} = \sigma(W^\top z_{\neg c} + b)$$

- Following the direction of W, for $\alpha \in [-10{,}10]$:

$$z_{\neg c}(\alpha) = z_{\neg c} + \alpha \frac{W}{\|W\|_2}$$



Some directions in the latent space encode undiscovered features:
- Parallel to base dimensions
- Aggregated from $z_{\neg c}$.

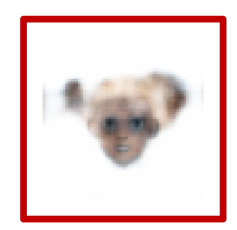
Enhancing Latent Disentanglement Through Contrastive Learning

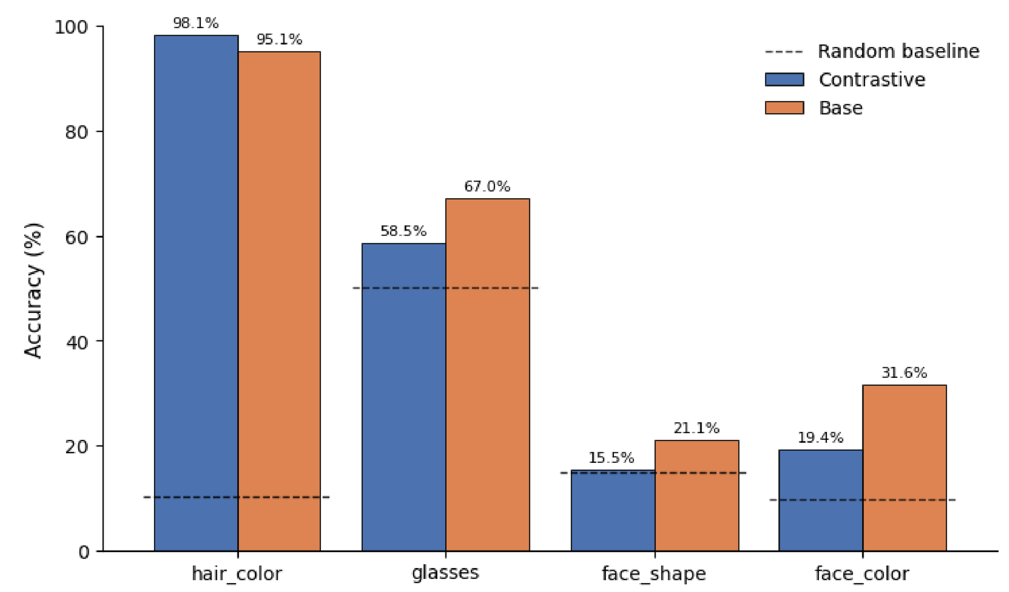
- Enforcing consistency in the representations of the latent label component z_c:
  - Images in the same class should have close representations.
  - Images in different classes should have distant representations.
- Gratifying **cosine similarity** between normalized latent encodings z_c.
- For an anchor i, a good $(z_c)_i$ should minimize:

$$\mathcal{L}_{\text{contrastive},i} = -\frac{1}{|P(i)|}\sum_{p\in P(i)} \log \frac{\exp\left(\mathbf{z}_{c,i}\cdot\mathbf{z}_{c,p}/\tau\right)}{\sum_{j\neq i}\exp\left(\mathbf{z}_{c,i}\cdot\mathbf{z}_{c,j}/\tau\right)},$$

where: $P(i) = \{p \in \mathcal{B} \mid p \neq i, y_p = y_i\}$.

- Upon unseen labels, the leakage ratio considerably decreases.
- For some unseen labels, z_c is as uninformative as noise.
- For the desired label, z_c is more informative.
- But !





Conclusion

- You can't eat the cake and keep it. Sorry !
- Future work may include introducing contrastiveness through a probabilistic modeling.

Reference:

- T. Joy, S. M. Schmon, P. H. S. Torr, N. Siddharth, and T. Rainforth. Capturing Label Characteristics in VAEs. In Proceedings of the International Conference on Learning Representations (ICLR), 2021.