SUPPLEMENTARY MATERIAL

Li Tan[1] *Hong Zhou*[2] *Fei Ye*[3,*]

[1]Adobe [2]Peking University [3]University of Electronic Science and Technology of China

Theorem 1. *Let $(\tilde{X}, \bar{X}, Y)$ be random variables and let g be a (possibly stochastic) encoder applied to both $\tilde{X}$ and $\bar{X}$. Assume the encoder is sufficient [1], then:*

$$I\big(g(\tilde{X}); Y \mid g(\bar{X})\big) = I\big(\tilde{X}; Y \mid \bar{X}\big). \tag{1}$$

Proof. Let $(\tilde{X}, \bar{X}, Y)$ be random variables and let g be a (possibly stochastic) *shared* encoder applied argument-wise to $\tilde{X}$ and $\bar{X}$. Assume *sufficiency*:

$$P(Y \mid \bar{X}) = P\big(Y \mid g(\bar{X})\big), \tag{S1}$$

$$P(Y \mid \tilde{X}, \bar{X}) = P\big(Y \mid g(\tilde{X}), g(\bar{X})\big). \tag{S2}$$

By the chain rule of conditional mutual information,

$$I(\tilde{X}; Y \mid \bar{X}) \;=\; H(Y \mid \bar{X}) - H(Y \mid \tilde{X}, \bar{X}). \tag{2}$$

By sufficiency (S1) and (S2),

$$H(Y \mid \bar{X}) = H\big(Y \mid g(\bar{X})\big), \tag{3}$$

$$H(Y \mid \tilde{X}, \bar{X}) = H\big(Y \mid g(\tilde{X}), g(\bar{X})\big). \tag{4}$$

Therefore,

$$I(\tilde{X}; Y \mid \bar{X}) = H\big(Y \mid g(\bar{X})\big) - H\big(Y \mid g(\tilde{X}), g(\bar{X})\big) \tag{5}$$

$$= I\big(g(\tilde{X}); Y \mid g(\bar{X})\big), \tag{6}$$

which proves the claim.

□

1. REFERENCES

[1] Yonglong Tian, Chen Sun, Ben Poole, Dilip Krishnan, Cordelia Schmid, and Phillip Isola, "What makes for good views for contrastive learning?," *Advances in neural information processing systems*, vol. 33, pp. 6827–6839, 2020.

Corresponding author: feiye@uestc.edu.cn