Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Supplementary materials for

Shaowu XU, Xibin JIA, Qianmei SUN, Jing CHANG, 2025. Temporal fidelity enhancement for video action recognition. *Front Inform Technol Electron Eng*, 26(8):1293-1304.
https://doi.org/10.1631/FITEE.2500164

Proof of Theorem 1

We follow that of Definition 1 of Pan et al. (2021) to prove Theorem 1 in the main text.

Let $\boldsymbol{z}_{\mathrm{N}}$ denote non-salient video embedding that captures action-irrelevant information and is semantically complementary to $\boldsymbol{z}_{\mathrm{S}}$. The objective function is formalized as:

$$\mathcal{L}_{\mathrm{DIB}}=-I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{y})+I(\boldsymbol{z}_{\mathrm{N}};\boldsymbol{y})+I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{z}_{\mathrm{N}}). \tag{S1}$$

Theorem 1 The DisenIB-based objective function, $\mathcal{L}_{\mathrm{DIB}}$, to be minimized is consistent with maximum compression.

Definition 1 (Consistency (Pan et al., 2021)) The lower-bounded cost functional $\mathcal{L}$ is consistent on maximum compression, if

$$\begin{aligned}&\forall\epsilon>0,\exists\delta>0,\quad \mathcal{L}-\mathcal{L}^*<\delta \implies\\ &|I(\boldsymbol{x};\boldsymbol{u})-H(\boldsymbol{y})|+|I(\boldsymbol{u};\boldsymbol{y})-H(\boldsymbol{y})|<\epsilon,\end{aligned} \tag{S2}$$

where $\mathcal{L}^*$ is the global minimum of $\mathcal{L}$.

Proof: The global minimum of $\mathcal{L}_{\mathrm{DIB}}$ is

$$\begin{aligned}\mathcal{L}^*_{\mathrm{DIB}}&=\min\mathcal{L}_{\mathrm{DIB}}\\ &=-\max I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{y})+\min I(\boldsymbol{z}_{\mathrm{N}};\boldsymbol{y})+\min I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{z}_{\mathrm{N}})\\ &=-H(\boldsymbol{y}).\end{aligned} \tag{S3}$$

We assume $\mathcal{L}_{\mathrm{DIB}}-\mathcal{L}^*_{\mathrm{DIB}}<\delta$, then we obtain the follows by combining Eq. (S1) and Eq. (S3):

$$H(\boldsymbol{y})-I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{y})<\delta,\quad I(\boldsymbol{z}_{\mathrm{N}};\boldsymbol{y})<\delta,\quad I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{z}_{\mathrm{N}})<\delta. \tag{S4}$$

Meanwhile, as $\boldsymbol{z}_{\mathrm{S}}$ and $\boldsymbol{z}_{\mathrm{N}}$ are semantically complementary, we can derive from Eq. (S4):

$$\begin{aligned}H(\boldsymbol{x})-I(\boldsymbol{x};\boldsymbol{z}_{\mathrm{N}},\boldsymbol{y})&=H(\boldsymbol{x}\mid\boldsymbol{z}_{\mathrm{N}},\boldsymbol{y})\\ &\le I(\boldsymbol{z}_{\mathrm{N}};\boldsymbol{y})+I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{z}_{\mathrm{N}})+H(\boldsymbol{y}\mid\boldsymbol{z}_{\mathrm{S}})\\ &<3\delta.\end{aligned} \tag{S5}$$

For given variables, we have Markov chains $\boldsymbol{z}_{\mathrm{S}}\leftrightarrow\boldsymbol{x}\leftrightarrow\boldsymbol{y}$, $\boldsymbol{z}_{\mathrm{N}}\leftrightarrow\boldsymbol{x}\leftrightarrow\boldsymbol{y}$, and $\boldsymbol{z}_{\mathrm{S}}\leftrightarrow\boldsymbol{x}\leftrightarrow\boldsymbol{z}_{\mathrm{N}}$. Since $\boldsymbol{x}$ contains all the information for deducing $\boldsymbol{y}$, we have

$$I(\boldsymbol{x};\boldsymbol{y})=H(\boldsymbol{y})-H(\boldsymbol{y}\mid\boldsymbol{x})=H(\boldsymbol{y}). \tag{S6}$$

According to the lemma proposed in Pan et al. (2021) about mutual information in Markov chains, we obtain

$$I(\boldsymbol{x};\boldsymbol{z}_{\mathrm{S}})+I(\boldsymbol{x};\boldsymbol{y})-I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{y})=I(\boldsymbol{x};\boldsymbol{z}_{\mathrm{S}},\boldsymbol{y}), \tag{S7}$$

$$I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{N}}) + I(\boldsymbol{x}; \boldsymbol{y}) - I(\boldsymbol{z}_{\mathrm{N}}; \boldsymbol{y}) = I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{N}}, \boldsymbol{y}), \tag{S8}$$

$$I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}) + I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{N}}) - I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{z}_{\mathrm{N}}) = I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}, \boldsymbol{z}_{\mathrm{N}}). \tag{S9}$$

By combining Eq. (S6) and Eq. (S7), and leveraging the inequality in Eq. (S4), we can obtain

$$\begin{aligned} I(\boldsymbol{x}; \boldsymbol{y}) - I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{y}) &= I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}, \boldsymbol{y}) - I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}) \\ &= H(\boldsymbol{y}) - I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{y}) \\ &< \delta. \end{aligned} \tag{S10}$$

By combining Eq. (S8) from Eq. (S9), and leveraging Eq. (S5), we can obtain

$$\begin{aligned} & I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}) - I(\boldsymbol{x}; \boldsymbol{y}) - I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}, \boldsymbol{z}_{\mathrm{N}}) + I(\boldsymbol{z}_{\mathrm{N}}; \boldsymbol{y}) \\ & = I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{z}_{\mathrm{N}}) - I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{N}}, \boldsymbol{y}) < 4\delta - H(\boldsymbol{x}). \end{aligned} \tag{S11}$$

By adding Eq. (S10) and Eq. (S11), and moving $H(\mathbf{x})$ from the right side to the left side, we have

$$H(\boldsymbol{x}) - I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}, \boldsymbol{z}_{\mathrm{N}}) + I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}, \boldsymbol{y}) - I(\boldsymbol{x}; \boldsymbol{y}) + I(\boldsymbol{z}_{\mathrm{N}}; \boldsymbol{y}) < 5\delta. \tag{S12}$$

According to the definition of mutual information, we have

$$\begin{aligned} H(\boldsymbol{x}) - I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}, \boldsymbol{z}_{\mathrm{N}}) &\geq 0, \\ I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}, \boldsymbol{y}) - I(\boldsymbol{x}; \boldsymbol{y}) &\geq 0, \\ I(\boldsymbol{z}_{\mathrm{N}}; \boldsymbol{y}) &\geq 0. \end{aligned} \tag{S13}$$

By combining Eq. (S12) and Eq. (S13), we further have

$$\begin{aligned} H(\boldsymbol{x}) - I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}, \boldsymbol{z}_{\mathrm{N}}) &\leq 5\delta, \\ I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}, \boldsymbol{y}) - I(\boldsymbol{x}; \boldsymbol{y}) &\leq 5\delta, \\ I(\boldsymbol{z}_{\mathrm{N}}; \boldsymbol{y}) &\leq 5\delta. \end{aligned} \tag{S14}$$

The data processing inequality (Cover & Thomas, 2012) indicates that the information loss is nonnegative. And we can obtain the upper bound of $I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}) - I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{y})$ by plugging $I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}, \boldsymbol{y}) - I(\boldsymbol{x}; \boldsymbol{y}) \leq 4\delta$ into Eq. (S7). Thus, we have

$$\begin{aligned} & 0 \leq I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}) - I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{y}) \leq 5\delta \iff \\ & |I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}) - I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{y})| \leq 5\delta. \end{aligned} \tag{S15}$$

On one hand, Definition 1 requires to find the upper and lower bound of $I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}) - H(\boldsymbol{y})$. By combining Eq. (S11) and Eq. (S5),

$$\begin{aligned} & I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}) - I(\boldsymbol{x}; \boldsymbol{y}) \\ & = I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{z}_{\mathrm{N}}) - I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{N}}, \boldsymbol{y}) + I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}, \boldsymbol{z}_{\mathrm{N}}) - I(\boldsymbol{z}_{\mathrm{N}}; \boldsymbol{y}) \\ & < 4\delta - H(\boldsymbol{x}) + I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}, \boldsymbol{z}_{\mathrm{N}}) - I(\boldsymbol{z}_{\mathrm{N}}; \boldsymbol{y}), \end{aligned} \tag{S16}$$

where $I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{z}_{\mathrm{N}}) - I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{N}}, \boldsymbol{y}) + I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}, \boldsymbol{z}_{\mathrm{N}}) - I(\boldsymbol{z}_{\mathrm{N}}; \boldsymbol{y}) \in (-10\delta, 4\delta)$ according to the inequality in Eq. (S14). Therefore, by plugging Eq. (S6) into Eq. (S16), we further have

$$|I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}) - I(\boldsymbol{x}; \boldsymbol{y})| = |I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}) - H(\boldsymbol{y})| \leq 10\delta. \tag{S17}$$

On the other hand, Definition 1 involves the determination of the upper and lower bound of $I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{y}) - H(\boldsymbol{y})$. To this end, we extend Eq. (S15) to include Eq. (S17) for estimating $I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{y})$ as follows:

$$\begin{aligned} & |I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{y}) - I(\boldsymbol{x}, \boldsymbol{y})| \\ & \leq |I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}) - I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{y})| + |I(\boldsymbol{x}; \boldsymbol{z}_{\mathrm{S}}) - I(\boldsymbol{x}; \boldsymbol{y})| \\ & < 15\delta, \end{aligned} \tag{S18}$$

By plugging Eq. (S6) into Eq. (S18) and being combined with Eq. (S17), we have

$$|I(\boldsymbol{x};\boldsymbol{z}_{\mathrm{S}})-H(\boldsymbol{y})|+|I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{y})-H(\boldsymbol{y})|<25\delta. \tag{S19}$$

As the above proof, $\forall\epsilon>0,\exists\delta=\epsilon/25>0$, they satisfy the follows:

$$\begin{aligned}&\mathcal{L}_{\mathrm{DIB}}-\mathcal{L}^{*}_{\mathrm{DIB}}<\delta \implies\\ &|I(\boldsymbol{x};\boldsymbol{z}_{\mathrm{S}})-H(\boldsymbol{y})|+|I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{y})-H(\boldsymbol{y})|<\epsilon,\end{aligned} \tag{S20}$$

which means that $\mathcal{L}_{\mathrm{DIB}}$ is consistent on maximum compression according to Definition 1.

Proof of Theorem 2

We follow Theorem 1 of Liang et al. (2020) to prove Theorem 2 in the main text.
Theorem 2 The global optimum for minimizing $\mathcal{L}_{\mathrm{DIB}}$ satisfies:

$$D^{*}=\arg\min_{D}\mathbb{E}\left[-\log p(\boldsymbol{y}|\boldsymbol{z}_{\mathrm{S}})+\log p(\boldsymbol{y}|\boldsymbol{z}_{\mathrm{N}})\right] \tag{S21}$$

where D^{*} denotes the optimal disentangler in Eq. (S1) that decomposes input segment embeddings $\boldsymbol{x}$ into salient and non-salient video embeddings, $\boldsymbol{z}_{\mathrm{S}}$ and $\boldsymbol{z}_{\mathrm{N}}$.

Proof sketch: We decompose $\mathcal{L}_{\mathrm{DIB}}$ into $\mathcal{L}_1=-I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{y})+I(\boldsymbol{z}_{\mathrm{N}};\boldsymbol{y})$ and $\mathcal{L}_2=I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{z}_{\mathrm{N}})$. First, we prove that D^{*} minimizes $\mathcal{L}_1$ by showing $E[-\log p(\boldsymbol{y}|\boldsymbol{z}_{\mathrm{S}})+\log p(\boldsymbol{y}|\boldsymbol{z}_{\mathrm{N}})]\geq E[-\log p(\boldsymbol{y}|\boldsymbol{z}^{*}_{\mathrm{S}})+\log p(\boldsymbol{y}|\boldsymbol{z}^{*}_{\mathrm{N}})]$ for any pair $(\boldsymbol{z}_{\mathrm{S}},\boldsymbol{z}_{\mathrm{N}})$, leading to $-I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{y})+I(\boldsymbol{z}_{\mathrm{N}};\boldsymbol{y})\geq -I(\boldsymbol{z}^{*}_{\mathrm{S}};\boldsymbol{y})+I(\boldsymbol{z}^{*}_{\mathrm{N}};\boldsymbol{y})$. Second, using a proof by contradiction, we demonstrate that minimizing $\mathcal{L}_1$ also minimizes $\mathcal{L}_2$, ensuring $I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{z}_{\mathrm{N}})$ is minimized. Thus, D^{*} provides the global optimum for minimizing $\mathcal{L}_{\mathrm{DIB}}$. Detailed proof can be found in .

Proof: Let $\mathcal{L}_{\mathrm{DIB}}=\mathcal{L}_1+\mathcal{L}_2$, where $\mathcal{L}_1=-I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{y})+I(\boldsymbol{z}_{\mathrm{N}};\boldsymbol{y})$ and $\mathcal{L}_2=I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{z}_{\mathrm{N}})$. Firstly, we prove that Eq. (S21) reaches the min $\mathcal{L}_1$. And then, we prove that $\mathcal{L}_2$ reaches the minimum while $\mathcal{L}_1$ has been minimized. Therefore, we can prove that Eq. (S21) is a global optimum of minimizing $\mathcal{L}_{\mathrm{DIB}}$.

(1) Given the definition of D^{*}, we have $D^{*}(\boldsymbol{x})=(\boldsymbol{z}^{*}_{\mathrm{S}},\boldsymbol{z}^{*}_{\mathrm{N}})$, and for any $\boldsymbol{z}_{\mathrm{S}}$ and $\boldsymbol{z}_{\mathrm{N}}$,

$$\begin{aligned}&E[-\log p(\boldsymbol{y}\mid\boldsymbol{z}_{\mathrm{S}})+\log p(\boldsymbol{y}\mid\boldsymbol{z}_{\mathrm{N}})]\\ &\geq E[-\log p(\boldsymbol{y}\mid\boldsymbol{z}^{*}_{\mathrm{S}})+\log p(\boldsymbol{y}\mid\boldsymbol{z}^{*}_{\mathrm{N}})].\end{aligned} \tag{S22}$$

As $\boldsymbol{y}$ is encoded from the labels, the value of $E[\log p(\boldsymbol{y})]$ and $E[\log p^{*}(\boldsymbol{y})]$ remain constant. By adding $E[\log p(\boldsymbol{y})]$ at both sides of Eq. (S22), we have

$$\begin{aligned}&E[\log p(\boldsymbol{y})]-E[\log p(\boldsymbol{y}\mid\boldsymbol{z}_{\mathrm{S}})]-\\ &E[\log p(\boldsymbol{y})]+E[\log p(\boldsymbol{y}\mid\boldsymbol{z}_{\mathrm{N}})]\\ &\geq E[\log p(\boldsymbol{y})]-E[\log p(\boldsymbol{y}\mid\boldsymbol{z}^{*}_{\mathrm{S}})]-\\ &E[\log p(\boldsymbol{y})]+E[\log p(\boldsymbol{y}\mid\boldsymbol{z}^{*}_{\mathrm{N}})].\end{aligned} \tag{S23}$$

According to the definition of mutual information, we can derive from Eq. (S23):

$$-I(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{y})+I(\boldsymbol{z}_{\mathrm{N}};\boldsymbol{y})\geq -I(\boldsymbol{z}^{*}_{\mathrm{S}};\boldsymbol{y})+I(\boldsymbol{z}^{*}_{\mathrm{N}};\boldsymbol{y}). \tag{S24}$$

Eq. (S24) indicates that D^{*} allows $\mathcal{L}_1$ to reach the minimum.

(2) To show that minimized $\mathcal{L}_1$ leads $\mathcal{L}_2$ to the minimum, we can use a proof by contradiction. Assume that while $\mathcal{L}_1$ reaches the minimum, there still exists D', satisfying that

$$I(\boldsymbol{z}^{*}_{\mathrm{S}};\boldsymbol{z}^{*}_{\mathrm{N}})-\min\mathcal{L}_2=I(\boldsymbol{z}^{*}_{\mathrm{S}};\boldsymbol{z}^{*}_{\mathrm{N}})-I(\boldsymbol{z}'_{\mathrm{S}};\boldsymbol{z}'_{\mathrm{N}})>0. \tag{S25}$$

Due to any pair $(\boldsymbol{z}_{\mathrm{S}};\boldsymbol{z}_{\mathrm{N}})$ is generated from mutually exclusive temporal attentions, we have $\boldsymbol{x}=\boldsymbol{z}_{\mathrm{S}}\cup\boldsymbol{z}_{\mathrm{N}}$. Under this premise, with there are Markov chains $\boldsymbol{z}_{\mathrm{S}}\leftrightarrow\boldsymbol{x}\leftrightarrow\boldsymbol{y}$, $\boldsymbol{z}_{\mathrm{N}}\leftrightarrow\boldsymbol{x}\leftrightarrow\boldsymbol{y}$, and $\boldsymbol{z}_{\mathrm{S}}\leftrightarrow\boldsymbol{x}\leftrightarrow\boldsymbol{z}_{\mathrm{N}}$, any

$\mathit{\Delta} = I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{z}_{\mathrm{N}}) - \min \mathcal{L}_2 \geq 0$ will lead to equal decrease in $I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{y})$ and increase in $I(\boldsymbol{z}_{\mathrm{N}}; \boldsymbol{y})$ by the same amount as $\mathit{\Delta}$. Therefore, according to Eq. (S26), we have

$$\begin{aligned} -I(\boldsymbol{z}'_{\mathrm{S}}; \boldsymbol{y}) + I(\boldsymbol{z}'_{\mathrm{N}}; \boldsymbol{y}) &= -\left(I(\boldsymbol{z}^*_{\mathrm{S}}; \boldsymbol{y}) + \mathit{\Delta}'\right) + I(\boldsymbol{z}^*_{\mathrm{N}}; \boldsymbol{y}) - \mathit{\Delta}' \\ &= -I(\boldsymbol{z}^*_{\mathrm{S}}; \boldsymbol{y}) + I(\boldsymbol{z}^*_{\mathrm{N}}; \boldsymbol{y}) - 2\mathit{\Delta}' \\ &< -I(\boldsymbol{z}^*_{\mathrm{S}}; \boldsymbol{y}) + I(\boldsymbol{z}^*_{\mathrm{N}}; \boldsymbol{y}), \end{aligned} \tag{S26}$$

where $\mathit{\Delta}' = I(\boldsymbol{z}^*_{\mathrm{S}}; \boldsymbol{z}^*_{\mathrm{N}}) - I(\boldsymbol{z}'_{\mathrm{S}}; \boldsymbol{z}'_{\mathrm{N}}) \geq 0$. As Eq. (S26) is a contradiction to the assumption that $\mathcal{L}_1$ reaches the minimum, it can be considered that minimized $\mathcal{L}_1$ leads $\mathcal{L}_2$ to the minimum.

As the above proof, Eq. (S21) explicitly minimizes $-I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{y}) + I(\boldsymbol{z}_{\mathrm{N}}; \boldsymbol{y})$ to its minimum value while implicitly reducing $I(\boldsymbol{z}_{\mathrm{S}}; \boldsymbol{z}_{\mathrm{N}})$ to its minimum value, which is a global optimum of minimizing the objective functional $\mathcal{L}_{\mathrm{DIB}}$.

References

Pan, Z., Niu, L., Zhang, J., & Zhang, L. (2021). Disentangled Information Bottleneck. Proceedings of the AAAI Conference on Artificial Intelligence, 35(10), 9285–9293.

Cover, T.M., & Thomas, J.A. (2012). *Elements of Information Theory.* Wiley.

Liang, J., Bai, B., Cao, Y., Bai, K., & Wang, F. (2020). Adversarial Infidelity Learning for Model Interpretation. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 286–296.