Towards an Incremental Unified Multimodal Anomaly Detection: Augmenting Multimodal Denoising From an Information Bottleneck Perspective (Supplementary Material)

Kaifang Long[1], Lianbo Ma[1*], Jiaqi Liu[2], Liming Liu[1], Guoyang Xie[3],
[1]Software College, Northeastern University, China, [2]UBC, Canada, [3]CATL, China
longkf@stumail.neu.edu.cn, malb@swc.neu.edu.cn, guoyang.xie@ieee.org

1. Overview

This supplementary material includes:

- Detailed theoretical analysis and proof related to information bottleneck regularization in our denoising framework of incremental unified multimodal anomaly detection. (Appendix A).
- P-AUROC and AUPRO scores on MVTec 3D-AD dataset (Appendix B).
- P-AUROC and AUPRO scores on Eyecndies dataset (Appendix C).
- A detailed description of the IB-IUMAD algorithm (see Algorithms 1 and Appendix D).

2. Appendix A: Theoretical Proofs

Let F_{fu}, F^g_{fu}, and Y be continuous random variables with corresponding support sets $\mathcal{F}$, $\mathcal{F}_g$, and Y, as well as probability distributions $P_{F_{fu}}$, $P_{F^g_{fu}}$, and P_Y. The mutual information between $P_{F^g_{fu}}$ and Y, along with its relationship to information entropy [3, 7], is defined as:

$$I(F^g_{fu};Y) \equiv \mathbb{E}\left[\log \frac{P_{F^g_{fu},Y}(F^g_{fu},Y)}{P_{F^g_{fu}}(F^g_{fu})P_Y(Y)}\right] = -\mathbb{E}[\log P_Y(Y)] + \mathbb{E}[\log P_{F^g_{fu},Y}(F^g_{fu},Y)] - \mathbb{E}[\log P_{F^g_{fu}}(F^g_{fu})] = -H(F^g_{fu},y) + H(F^g_{fu}) + H(Y) = H(F^g_{fu}) - H(F^g_{fu}|Y).$$

Additionally, given Y, the conditional mutual information of $P_{F_{fu}}$ and $P_{F^g_{fu}}$ is defined as:

$$I(F_{fu};F^g_{fu}|Y) \equiv \mathbb{E}\left[\log \frac{P_{F_{fu},F^g_{fu}|Y}(F_{fu},F^g_{fu}|Y)}{P_{F_{fu},Y}(F_{fu},Y)P_{F^g_{fu},Y}(F^g_{fu},Y)}\right] = \int_{\mathcal{F},\mathcal{F}_g,Y} P_{F_{fu},F^g_{fu}|Y}(F_{fu},F^g_{fu}|Y)P_Y(Y) \log \frac{P_{F_{fu},F^g_{fu}|Y}(F_{fu},F^g_{fu}|Y)}{P_{F_{fu},Y}(F_{fu},Y)P_{F^g_{fu},Y}(F^g_{fu},Y)} dF_{fu}dF^g_{fu}dY.$$

Based on the aforementioned definitions of mutual information, conditional mutual information, and entropy [4–6], we derive the following corollaries:

Corollary 1. *Given a random variable Y, the mutual information between features F_{fu} and F^g_{fu}, can be equivalently rewritten as $I(F_{fu};F^g_{fu}) = I(F_{fu};F^g_{fu}|Y) + I(F^g_{fu};Y)$ according to the chain rule.*

Proof. Based on mutual information theory, we first know that $I(F^g_{fu};Y) = -\mathbb{E}[\log P_Y(Y)] + \mathbb{E}[\log P_{F^g_{fu},Y}(F^g_{fu},Y)] - \mathbb{E}[\log P_{F^g_{fu}}(F^g_{fu})] = H(F^g_{fu}) - H(F^g_{fu}|Y)$. Then, since F^g_{fu} is derived from F_{fu}, it follows that $P_{F_{fu},F^g_{fu}}(F_{fu},F^g_{fu}) = P_{F_{fu}}(F_{fu})$, implying $I(F_{fu};F^g_{fu}) = \mathbb{E}\ [-\log\ P_{F^g_{fu}}(F^g_{fu})] = [\log \frac{P_{F_{fu},F^g_{fu}}(F_{fu},F^g_{fu})}{P_{F^g_{fu}}(F^g_{fu})P_{F_{fu}}(F_{fu})}] = H(F^g_{fu})$. Similarly, we can deduce that $I(F_{fu},F^g_{fu}|Y) = H(F^g_{fu}|Y)$. Notably, given that mutual information is symmetric, it follows that $I(F_{fu};Y) = I(Y;F_{fu})$ and $I(F^g_{fu};Y) = I(Y;F^g_{fu})$. Combining these relations, we can prove that $I(F_{fu};F^g_{fu}) = I(F_{fu};F^g_{fu}|Y) + I(F^g_{fu};Y)$.

Corollary 2. *If the KL divergence between $p(Y|F_{fu})$ and $p(Y|F^g_{fu})$ is 0, that is, $\mathrm{KL}[P(Y|F_{fu})||P(Y|F^g_{fu})] = 0$, we have $I(Y;F_{fu}) - I(Y;F^g_{fu}) = I(F_{fu};Y) - I(F^g_{fu};Y) = 0$.*

*Corresponding author.

Table 1. AUPRO scores on MVTec 3D-AD dataset (10-0 with 0 step). The red / blue indicates the best/second-best results.

	Method	Year	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
RGB	UniAD	NIPS22	84.4	**96.3**	93.4	**88.7**	**96.1**	55.8	**90.4**	91.1	**94.3**	**90.6**	88.1
	SimpleNet	CVPR23	70.4	86.8	84.4	66.6	83.0	66.7	74.8	72.8	92.8	77.9	77.6
	DeSTSeg	CVPR23	77.6	64.1	14.2	40.9	31.2	63.6	48.2	6.2	90.4	27.3	46.4
	DiAD	AAAI24	**93.8**	94.5	**94.6**	83.5	89.6	**69.1**	**94.2**	**93.9**	**96.5**	68.8	87.8
	IUF	ECCV24	85.8	93.4	**93.8**	86.5	93.7	68.4	89.6	88.6	93.4	**92.2**	**88.5**
	CDAD	CVPR25	87.1	95.2	88.3	77.3	95.8	**72.0**	88.9	88.2	92.4	90.0	87.5
	IB-IUMAD	-	**89.1**	**97.0**	91.8	**91.3**	**95.9**	64.3	89.9	**92.5**	93.5	87.6	**89.3**
3D	DiAD	AAAI24	61.5	62.8	70.4	**63.7**	**76.5**	35.7	**68.8**	76.7	65.7	55.6	63.7
	IUF	ECCV24	**64.6**	**64.2**	**81.3**	**60.6**	**76.5**	26.8	**71.7**	**76.8**	**68.8**	**62.6**	**65.4**
	CDAD	CVPR25	39.2	48.7	36.8	33.7	45.2	**40.4**	39.8	36.1	44.9	40.7	40.6
	IB-IUMAD	-	**68.9**	**63.6**	**81.1**	58.3	**82.9**	**39.1**	66.8	**83.0**	**69.4**	**62.5**	**67.6**
RGB+3D	DiAD	AAAI24	**95.2**	94.8	92.9	85.6	90.7	65.3	**93.5**	**94.6**	95.1	71.2	87.9
	IUF	ECCV24	87.5	**96.6**	**95.3**	**90.3**	**96.2**	64.5	89.8	88.7	**97.3**	85.4	**89.2**
	CDAD	CVPR25	87.8	**96.1**	87.6	79.4	**96.1**	**74.3**	87.2	86.9	94.3	**91.2**	88.1
	IB-IUMAD	-	**92.6**	90.7	**95.4**	**90.5**	95.3	**68.9**	**92.5**	**92.2**	**97.4**	**88.7**	**90.4**

Proof.

$$
\begin{aligned}
& I(F_{fu};Y) - I(F_{fu}^g;Y) = I(Y;F_{fu}) - I(Y;F_{fu}^g) \\
&= -\iint P(F_{fu}^g)P(Y|F_{fu}^g)\log P(Y|F_{fu}^g)dF_{fu}^g dY \\
&+ \iint P(F_{fu})P(Y|F_{fu})\log P(Y|F_{fu})dF_{fu}dY \\
&= \iint P(F_{fu})P(Y|F_{fu})\log\left[\frac{P(Y|F_{fu})}{P(Y|F_{fu}^g)}P(Y|F_{fu}^g)\right] \\
&\quad dF_{fu}dY \\
&- \iint P(F_{fu}^g)P(Y|F_{fu}^g)\log\left[\frac{P(Y|F_{fu}^g)}{P(Y|F_{fu})}P(Y|F_{fu})\right] \\
&\quad dF_{fu}^g dY \\
&= -\int P(F_{fu}^g)\mathrm{KL}\left[P(Y|F_{fu}^g)||P(Y|F_{fu})\right]dF_{fu}^g \\
&- \iint P(F_{fu}^g)P(Y|F_{fu}^g)\log P(Y|F_{fu})dF_{fu}^g dY \\
&+ \int P(F_{fu})\mathrm{KL}\left[P(Y|F_{fu})||P(Y|F_{fu}^g)\right]dF_{fu} \\
&- \iint P(F_{fu})P(Y|F_{fu})\log P(Y|F_{fu}^g)dF_{fu}dY \\
&= \mathbb{E}_{F_{fu}}\left[\mathrm{KL}[P(Y|F_{fu})||P(Y|F_{fu}^g)]\right] \\
&- \mathbb{E}_{F_{fu}^g}\left[\mathrm{KL}[P(Y|F_{fu}^g)||P(Y|F_{fu})]\right] \\
&+ \int P(Y)\log\frac{P(Y|F_{fu}^g)}{p(Y|F_{fu})}dY \\
&\le \mathbb{E}_{F_{fu}}\left[\mathrm{KL}[P(Y|F_{fu})||P(Y|F_{fu}^g)]\right] \\
&+ \int P(Y)\log\frac{P(Y|F_{fu}^g)}{p(Y|F_{fu})}dY
\end{aligned}
$$

By Jensen's inequality and the strict convexity of $-\log$, we conclude that the KL divergence is non-negative [3]. When $\mathrm{KL}[P(Y|F_{fu})||P(Y|F_{fu}^g)] = 0$, it follows that $P(Y|F_{fu}^g) = P(Y|F_{fu})$ almost everywhere, which means that $\int P(Y)\log\frac{P(Y|F_{fu}^g)}{p(Y|F_{fu})}dY = 0$, and we have $I(Y|F_{fu}) - I(Y|F_{fu}^g) \le 0$. Therefore, combining these, we prove that when the KL divergence between $p(Y|F_{fu})$ and $p(Y|F_{fu}^g)$ is 0, i.e., $\mathrm{KL}[P(Y|F_{fu})||P(Y|F_{fu}^g)] = 0$, we have $I(Y;F_{fu}) - I(Y;F_{fu}^g) = 0$.

To this end, based on ***Corollary* 1** and ***Corollary* 2**, we conclude that using KL divergence as the target loss function between F_{fu} and F_{fu}^g effectively eliminates redundant information from the fused multimodal features.

3. Appendix B: AUPRO and P-AUROC on MVTec 3D-AD

To validate the effectiveness of the proposed method, we report the P-AUROC and AUPRO metrics of IB-IUMAD under the setting of 10-0 with 0 step. As shown in Tables 1 and 2, on the MVTec 3D-AD dataset [1], IB-IUMAD consistently outperforms the state-of-the-art unified multimodal anomaly detection approaches in most test cases. Specifically, when trained solely with RGB images, IB-IUMAD achieves improvements of 3.3% and 11.7% over SimpleNet in terms of P-AUROC and AUPRO metrics, respectively (0.4% and 1.2% higher than UniAD, and 0.5% and 0.8% higher than IUF, respectively). When both RGB and 3D depth images are used for training, IB-IUMAD achieves 0.6% and 1.2% higher than IUF in terms of P-AUROC and AUPRO metrics, respectively (0.4% and 2.5% higher than DiAD, and 0.4% and 2.3% higher than CDAD, respectively). In fact, the abovementioned performance enhancements exceed state-of-the-art (SOTA) methods. These experimental results validate the critical role of eliminating spurious and redundant features in improving the perfor-

Table 2. P-AUROC scores on MVTec 3D-AD dataset (10-0 with 0 step). The red / blue indicates the best/second-best results.

	Method	Year	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
RGB	UniAD	NIPS22	97.6	**98.9**	98.0	**97.5**	**99.1**	82.2	97.4	97.6	99.0	98.0	96.5
	SimpleNet	CVPR23	93.2	95.2	96.4	90.5	95.3	87.8	92.9	91.0	99.3	93.8	93.6
	DeSTSeg	CVPR23	**98.7**	97.8	86.9	93.3	97.3	**95.7**	95.9	89.2	98.8	97.0	95.1
	DiAD	AAAI24	98.5	98.4	**98.6**	94.3	97.2	89.8	**98.4**	98.0	99.3	91.8	96.4
	IUF	ECCV24	97.9	98.2	98.5	97.0	98.9	81.8	97.4	96.6	**99.4**	98.1	96.4
	CDAD	CVPR25	97.5	**98.9**	95.8	95.1	**99.1**	92.3	96.9	96.0	99.0	97.7	96.8
	IB-IUMAD	-	97.4	97.1	97.4	**97.5**	98.6	88.7	96.6	**98.1**	99.2	**98.5**	**96.9**
3D	DiAD	AAAI24	88.3	90.6	94.4	84.9	**95.1**	62.8	91.3	**94.5**	91.6	**88.5**	88.2
	IUF	ECCV24	**90.3**	90.0	94.3	86.8	93.1	60.1	90.2	93.6	90.8	84.5	87.4
	CDAD	CVPR25	70.6	75.8	71.4	71.6	76.0	**77.7**	75.8	73.1	75.0	76.8	74.4
	IB-IUMAD	-	89.3	**90.8**	**94.7**	**87.4**	93.8	64.1	**91.7**	93.4	**91.9**	87.8	**88.5**
RGB+3D	DiAD	AAAI24	97.6	96.8	98.1	**97.4**	98.6	90.4	**97.7**	97.6	**99.5**	97.8	97.2
	IUF	ECCV24	97.8	97.8	97.7	96.7	**99.3**	90.1	97.5	96.7	99.4	97.2	97.0
	CDAD	CVPR25	**98.4**	**98.8**	96.1	95.5	99.1	93.4	97.1	96.6	98.7	**97.9**	97.2
	IB-IUMAD	-	97.9	97.6	**99.0**	**97.4**	99.2	**93.5**	97.6	**97.7**	99.4	96.4	**97.6**

Table 3. AUPRO scores on Eyecandies dataset (10-0 with 0 step). The red / blue indicates the best/second-best results.

	Method	Year	Candy Cane	Chocolate Cookie	Chocolate Praline	Confetto	Gummy Bear	Hazelnut Truffle	Licorice Sandwish	Lollipop	Marsh-mallow	Peppermint Candy	Mean
RGB	DiAD	AAAI24	86.8	83.8	76.7	96.5	70.5	53.2	85.1	88.7	93.1	93.8	82.8
	IUF	ECCV24	87.8	87.9	77.8	96.6	72.0	52.7	86.7	89.2	94.6	94.3	84.0
	CDAD	CVPR25	80.7	86.2	69.8	96.4	72.8	**66.4**	79.9	**91.4**	86.4	91.3	82.1
	IB-IUMAD	-	**89.8**	**89.5**	**80.3**	**97.5**	**74.5**	53.2	**88.2**	90.3	**95.0**	**95.0**	**85.3**
3D	DiAD	AAAI24	62.1	29.3	34.9	**50.3**	22.4	**21.3**	**43.2**	56.6	39.8	54.3	41.4
	IUF	ECCV24	79.6	28.7	26.7	44.9	18.3	3.2	41.9	54.2	42.8	53.8	39.4
	CDAD	CVPR25	48.3	13.6	**38.3**	20.1	7.2	17.8	32.6	51.2	40.1	33.2	30.2
	IB-IUMAD	-	**82.5**	**30.9**	31.8	49.6	**25.7**	10.4	42.5	**59.8**	**46.3**	**57.7**	**43.7**
RGB + 3D	DiAD	AAAI24	88.7	85.1	75.6	94.3	73.8	56.2	**89.2**	92.3	93.4	**95.7**	84.4
	IUF	ECCV24	**90.1**	89.7	**81.2**	96.4	74.7	54.5	88.6	90.1	**94.9**	95.2	85.5
	CDAD	CVPR25	86.9	85.4	69.7	96.6	69.5	**68.6**	83.7	**92.8**	90.1	93.2	83.7
	IB-IUMAD	-	89.5	**90.4**	80.9	**97.4**	**75.2**	58.7	87.9	91.1	94.6	**95.7**	**86.1**

Table 4. P-AUROC scores on Eyecandies dataset (10-0 with 0 step). The red / blue indicates the best/second-best results.

	Method	Year	Candy Cane	Chocolate Cookie	Chocolate Praline	Confetto	Gummy Bear	Hazelnut Truffle	Licorice Sandwish	Lollipop	Marsh-mallow	Peppermint Candy	Mean
RGB	DiAD	AAAI24	95.8	96.2	91.4	97.9	89.5	87.8	95.5	97.3	98.9	96.7	94.7
	IUF	ECCV24	96.1	95.8	**92.8**	98.6	**91.1**	86.4	95.7	96.9	98.8	97.5	95.0
	CDAD	CVPR25	94.8	95.6	92.0	99.3	87.1	**88.9**	94.6	96.6	97.6	**98.5**	94.5
	IB-IUMAD	-	**96.2**	**97.7**	92.5	**99.4**	89.8	86.1	**96.4**	**97.4**	**99.3**	98.1	**95.3**
3D	DiAD	AAAI24	94.9	65.1	60.5	78.8	56.8	60.6	80.1	85.1	74.8	79.2	73.6
	IUF	ECCV24	**95.0**	**69.2**	59.8	**81.8**	56.2	59.4	80.7	**87.6**	76.0	80.7	74.6
	CDAD	CVPR25	78.9	54.5	56.9	60.4	58.3	61.4	62.3	74.1	58.7	56.8	62.2
	IB-IUMAD	-	**95.0**	65.8	**64.7**	78.6	**62.3**	**62.7**	**81.9**	85.5	**76.7**	**81.6**	**75.5**
RGB + 3D	DiAD	AAAI24	**98.6**	**97.9**	92.3	99.3	89.6	85.7	95.8	**98.1**	**97.8**	97.1	95.2
	IUF	ECCV24	96.5	97.3	92.7	99.3	90.8	88.0	96.3	97.1	99.2	98.3	95.6
	CDAD	CVPR25	94.6	96.5	92.4	**99.5**	90.1	**89.1**	94.2	97.3	97.2	97.6	94.9
	IB-IUMAD	-	97.1	97.6	**94.7**	**99.5**	**93.8**	87.9	**96.6**	96.7	99.2	**98.6**	**96.2**

mance of incremental unified anomaly detection tasks.

4. Appendix C: AUPRO and P-AUROC on Eyecandies.

Tables 3 and 4 demonstrate the AUPRO and P-AUPRO scores under the setting of 10-0 with 0 step on the Eyecandies dataset [2]. It is clear that IB-IUMAD consistently outperforms the baselines in most cases. In particular, when trained with both RGB and depth images, IB-IUMAD achieves 0.6% and 0.6% improvements compared to IUF in terms of P-AUROC and AUPRO scores, respectively (1.0% and 1.7% higher than DiAD, and 1.3% and 2.4% superior to CDAD, respectively). These experimental results again demonstrate that the proposed denoising framework not only effectively eliminates inter-object spurious feature interference, but also filters out redundant information from the fused multimodal features.

5. Appendix D: Algorithms

We provide a detailed description of the IB-IUMAD implementation algorithm. Taking the MVTec 3D-AD dataset as an example, the training process primarily consists of two stages: first, we train a base model across six object categories; then, the remaining four objects are introduced sequentially through four incremental learning steps (i.e., 6-1 with 4 steps).

For the basic model training stage, we first employ the multimodal feature extraction network (Φ_{MFEN}) to extract the features of RGB (I_{rgb}) and depth (I_{depth}) images, respectively, and then generate abnormal RGB (A_{rgb}) and depth (A_{depth}) features through feature jitter. Subsequently, the Mamba decoder is used to extract fine-grained features (i.e., M_{rgb} and M_{depth}) from I_{rgb} and I_{depth}, aiming to introduce label information of the object category to mitigate interference from inter-object features. Next, the extracted features A_{rgb}, A_{depth}, M_{rgb}, and M_{depth} are fed into the multimodal reconstruction network (Φ_{MRN}) for feature reconstruction. Finally, we utilize the cross-attention mechanism to fuse the reconstructed features R_r and R_d, and adopt information bottleneck regularization to filter out noise features in F_{fusion}, thereby obtaining the final multimodal fusion feature F^g_{fusion}. The obtained F^g_{fusion} is fed into the discriminator (Φ_{dis}) for anomaly score discrimination. For the incremental training phase, the training process is similar to the basic model training. Notably, in Algorithm 1, $I \in [0, 4]$ represents the incremental step, M_{ask} indicates the ground-truth anomaly segmentation images, and Y denotes the ground-truth label of objects.

Algorithm 1 IB-IUMAD pseudo-code.

Input: Basic training data ($\mathcal{D}^B_{trian}$), Incremental training data ($\mathcal{D}^I_{trian}$), Test data ($\mathcal{D}_{test}$).
Output: Trained IUMAD model.
/*Basic model training stage*/
for $e \leftarrow 0$ **to** Epochs **do**
  for $I_{rgb}, I_{depth}, M_{ask} \leftarrow \mathcal{D}^B_{trian}$ **do**
    $A_{rgb}, A_{depth} = \Phi_{MFEM}(I_{rgb}, I_{depth})$
    $M_{rgb}, M_{depth} = \Phi_{Mamba}(I_{rgb}, I_{depth})$
    $R_r, R_d = \Phi_{MRN}(A_{rgb}, M_{rgb}, A_{depth}, M_{depth})$
    $F_{fuison} = Cross\_attention(R_r, R_d)$
    $F^g_{fuison} = \Phi_{IB}(F_{fuison})$
    $M = \Phi_{Dis}(I_{rgb}, F^g_{fuison},)$
    $\mathcal{L}_{rgb} = \Phi_{CE\_loss}(M_{rgb}, Y)$
    $\mathcal{L}_{Depth} = \Phi_{CE\_loss}(M_{Depth}, Y)$
    $\mathcal{L}_{IB} = \Phi_{KL\_loss}(Y_{F^g_{fuison}}, Y_{F_{fuison}})$
    $\mathcal{L}_{fuison} = \Phi_{MSE\_loss}(I_{rgb}, F^g_{fuison}, M_{ask})$
    $\mathcal{L}_{total} = \mathcal{L}_{rgb} + \mathcal{L}_{Depth} + \mathcal{L}_{IB} + \mathcal{L}_{fuison}$
  end for
end for
/*Incremental model training phase*/
for $I \leftarrow 0$ **to** Epochs **do**
  for $e \leftarrow 0$ **to** Epochs **do**
    for $I_{rgb}, I_{depth}, M_{ask} \leftarrow \mathcal{D}^I_{trian}$ **do**
      $A_{rgb}, A_{depth} = \Phi_{MFEM}(I_{rgb}, I_{depth})$
      $M_{rgb}, M_{depth} = \Phi_{Mamba}(I_{rgb}, I_{depth})$
      $R_r, R_d = \Phi_{MRN}(A_{rgb}, M_{rgb}, A_{depth}, M_{depth})$
      $F_{fuison} = Cross\_attention(R_r, R_d)$
      $F^g_{fuison} = \Phi_{IB}(F_{fuison})$
      $M = \Phi_{Dis}(I_{rgb}, F^g_{fuison})$
      $\mathcal{L}_{rgb} = \Phi_{CE\_loss}(M_{rgb}, Y)$
      $\mathcal{L}_{Depth} = \Phi_{CE\_loss}(M_{Depth}, Y)$
      $\mathcal{L}_{IB} = \Phi_{KL\_loss}(Y_{F^g_{fuison}}, Y_{F_{fuison}})$
      $\mathcal{L}_{fuison} = \Phi_{MSE\_loss}(I_{rgb}, F^g_{fuison}, M_{ask})$
      $\mathcal{L}_{total} = \mathcal{L}_{rgb} + \mathcal{L}_{Depth} + \mathcal{L}_{IB} + \mathcal{L}_{fuison}$
    end for
  end for
end for

References

[1] Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. In *VISIGRAPP*, 2022. 2

[2] Luca Bonfiglioli, Marco Toschi, Davide Silvestri, Nicola Fioraio, and Daniele De Gregorio. The eyecandies dataset for unsupervised multimodal anomaly detection and localization. In *Proceedings of the ACCV*, 2022. 3

[3] Yingying Fang, Shuang Wu, Sheng Zhang, Chaoyan Huang, Tieyong Zeng, Xiaodan Xing, Simon Walsh, and Guang Yang. Dynamic multimodal information bottleneck for multimodality classification. In *WACV*, 2024. 1, 2

[4] Marco Federici, Anjan Dutta, Patrick Forré, Nate Kushman, and Zeynep Akata. Learning robust representations via multi-view information bottleneck. In *ICLR*, 2020. 1

[5] Zhenguang Liu, Runyang Feng, Haoming Chen, Shuang Wu, Yixing Gao, Yunjun Gao, and Xiang Wang. Temporal feature alignment and mutual information maximization for video-based human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11006–11016, 2022.

[6] Xudong Tian, Zhizhong Zhang, Shaohui Lin, Yanyun Qu, Yuan Xie, and Lizhuang Ma. Farewell to mutual information: Variational distillation for cross-modal person re-identification. In *Proceedings of the CVPR*, 2021. 1

[7] Ruiwen Yuan, Yongqiang Tang, Yanghao Xiao, and Wensheng Zhang. Ibcs: Learning information bottleneck-constrained denoised causal subgraph for graph classification. *IEEE TPAMI*, 2024. 1