Revisiting Multimodal Fusion for 3D Anomaly Detection from An Architectural Perspective (Supplementary Material)

Anonymous submission

Overview

This supplementary material includes:

- Details of theoretical analysis and proof of the impact of inter- and intra-module fusion design on 3D-AD performance (Section A);
- Details of the examples of the multimodal fusion types (Section B);
- Details about optimal connections (identified by α^{in}) between input nodes and intermediate nodes in each cell of MSMs (Section C);
- The complete search process of 3D-ADNAS and the pseudo-code of inter- and intra-module fusion (see Algorithms 1 and 2);
- P-AUROC scores on Eyecandies dataset (Section D);
- AUPRO scores on MVTec 3D-AD dateset (Section E);
- Detailed results of few-shot setting on Eyecandies and MVTec 3D-AD datasets (Section F);
- Visualization results of all categories on the MVTec 3D-AD dataset. (Section G);

A. Theoretical Proofs

According to the Dempster-Shafer Evidence Theory (DST) (Han et al. 2021), we observe that the belief mass (b) and uncertainty (u) (Liu et al. 2017; Han et al. 2022; Xu et al. 2022) play a critical role in evaluating the trustworthiness of the predictive results of the target model. Inspired by this, we can employ the principle of DST to explore the impact of target intra-module and inter-module fusion design on 3D-AD. In the following, we provide the details of theoretical proofs involved in the section of the **Revisiting 3D-AD Fusion Architecture** in the main paper. At first, we consider the impact of inter-module fusion design, as follows.

For the N-class classification task of MSM, suppose that the category probabilities for each MSM can be obtained through the Variational Dirichlet, e.g., the early MSM a is $e_a = [e_a^1, e_a^2, e_a^3, ..., e_a^N]$, the middle MSM m is $e_m = [e_m^1, e_m^2, e_m^3, ..., e_m^N]$, and the late MSM l is $e_l = [e_l^1, e_l^2, e_l^3, ..., e_l^N]$ (Bishop and Nasrabadi 2006; Kingma and Welling 2014). Then, the DST is utilized to acquire the opinion of each MSM, which each opinion comprises of belief masses (b) and uncertainty (u), e.g., $\{\{b_a^n\}_{n=1}^N, u_a\}$ for early MSM opinion A, $\{\{b_m^n\}_{n=1}^N, u_m\}$ for middle MSM opinion M, and $\{\{b_l^n\}_{n=1}^N, u_l\}$ for late MSM opinion L. For the late MSM opinion L, its belief mass b_l^n and uncertainty u_l can be obtained as:

$$b_l^n = \frac{e_l^n}{S_l}, \quad u_l = \frac{N}{S_l}, \tag{1}$$

$$u_l + \sum_{n=1}^{N} b_l^n = 1, \tag{2}$$

$$S_l = (e_l^1 + 1) + (e_l^2 + 1) + ... + (e_l^N + 1), \tag{3}$$

where S_l is the Dirichlet strength, e_l^n is the category probability of late, b_l^n is the belief mass of the n-th class of the late MSM L (a larger b_l^n will lead to a larger e_l^n), u_l represents the uncertainty of late MSM (a larger u_l will reduce the trustworthiness of the classification result). Similarly, for other MSMs (A and E), their corresponding opinions can be obtained using the same way as used for the late MSM.

Given the opinion of a specific MSM (e.g., $L = \{\{b_l^n\}_{n=1}^N, u_l\}$), we aim to theoretically analyze whether the fusion of additional opinion (e.g., the middle MSM: $M = \{\{b_m^n\}_{n=1}^N, u_m\}$) will influence the model's classification accuracy. Then, following the combination rule of DST, we fuse L into M and form a new opinion $F = \{\{b_f^n\}_{n=1}^N, u_f\}$. Here, b_f^n and u_f are new belief mass and uncertainty, respectively, which are given by

$$b_f^n = \frac{1}{1-z}(b_l^n b_m^n + b_m^n u_l + b_l^n u_m), \tag{4}$$

$$u_f = \frac{1}{1-z}(u_l u_m), \tag{5}$$

where $z = \sum_{i \neq j} b_l^i b_m^j$ ($i, j \in [1, 2, ..., N]$) is the measure of the conflict quantity (as shown in the white block of Fig. 1) between two belief mass sets of L and M, and $\frac{1}{1-z}$ is used as the normalization factor.

Then, we can give the following propositions.

***Proposition* 1.** Under the conditions $b_m^g \geq b_l^{max}$, where $g \in N$ is the index of the ground-truth label, and b_l^{max} is the largest in $\{b_l^n\}_{n=1}^N$, fusing another opinion M makes the new opinion F satisfy $b_f^g \geq b_l^g$.

Proof.

$$
\begin{aligned}
b_f^g &= \frac{b_l^g b_m^g + b_l^g u_m + b_m^g u_l}{1-z} \\
&= \frac{b_l^g b_m^g + b_l^g u_m + b_m^g u_l}{\sum_{n=1}^N b_m^n b_l^n + u_m + u_l - u_m u_l} \\
&\geq \frac{b_l^g b_m^g + b_l^g u_m + b_l^{max} u_l}{\sum_{n=1}^N b_m^n b_l^n + u_m + u_l - u_m u_l} \\
&\geq \frac{b_l^g b_m^g + b_l^g u_m + b_l^{max} u_l}{\sum_{n=1}^N b_m^n b_l^{max} + u_m + u_l - u_m u_l} \\
&\geq \frac{b_l^g b_m^g + b_l^g u_m + b_l^g u_l}{b_l^{max} \sum_{n=1}^N b_m^n + u_m + u_l - u_m u_l} \\
&\geq \frac{b_l^g (b_m^g + u_m + u_l)}{b_l^{max}(1-u_m) + u_m + u_l - u_m u_l} \\
&\geq \frac{b_l^g (b_m^g + u_m + u_l)}{b_l^{max} + u_l + u_m} \\
&\geq b_l^g \frac{(b_l^{max} + u_m + u_l)}{b_l^{max} + u_m + u_l} \\
&\geq b_l^g,
\end{aligned}
$$

where, $1-z$ is derived by

$$
\begin{aligned}
1 - z &= u_m \sum_{n=1}^N b_l^n + u_l \sum_{n=1}^N b_m^n + u_m u_l + \sum_{n=1}^N b_m^n b_l^n \\
&= \sum_{n=1}^N b_m^n b_l^n + u_m(1-u_l) + u_l(1-u_m) + u_m u_l \\
&= \sum_{n=1}^N b_m^n b_l^n + u_m - u_m u_l + u_l - u_m u_l + u_m u_l \\
&= \sum_{n=1}^N b_m^n b_l^n + u_m + u_l - u_m u_l.
\end{aligned}
$$

***Proposition* 2.** When u_m is large, $b_l^g - b_f^g$ will be limited, and it will have a negative correlation with u_m. As a special case, when u_m is large enough (i.e., u_m = 1), fusing another opinion will not reduce the performance (i.e., $b_f^g = b_l^g$).

Proof.

$$
b_l^g - b_f^g = b_l^g - \frac{b_l^g b_m^g + b_l^g u_m + b_m^g u_l}{1-z}
$$

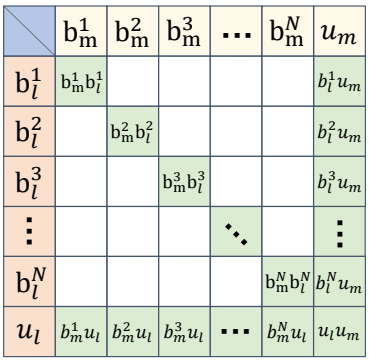

Figure 1: The combination rule of DST. Given the middle MSM opinion M (yellow block) and the late MSM opinion L (orange block), we combine them to form a new opinion F (green block). The white block is the measure of the conflict quantity between two belief mass sets of L and M.

$$
\begin{aligned}
&= b_l^g - \frac{b_l^g b_m^g + b_l^g u_m + b_m^g u_l}{\sum_{n=1}^N b_m^n b_l^n + u_m + u_l - u_m u_l} \\
&\leq b_l^g - \frac{b_l^g b_m^g + b_l^g u_m + b_m^g u_l}{b_m^{max} \sum_{n=1}^N b_l^n + u_m + u_l - u_m u_l} \\
&\leq b_l^g - \frac{b_l^g u_m}{b_m^{max} \cdot 1 + u_m + u_l - u_m u_l} \\
&\leq b_l^g - \frac{b_l^g u_m}{1 + u_l - u_m u_l},
\end{aligned}
$$

Then, we continue to simplify the above formula to make it more concise, as follows:

$$
\begin{aligned}
b_l^g - \frac{b_l^g u_m}{1 + u_l - u_m u_l} &= b_l^g (1 - \frac{u_m}{1 + u_l - u_l u_m}) \\
&= b_l^g \frac{1 + u_l - u_l u_m - u_m}{1 + u_l - u_l u_m} \\
&= b_l^g \frac{(1+u_l)(1-u_m)}{1 + u_l(1-u_m)} \\
&= b_l^g \frac{1+u_l}{\frac{1}{(1-u_m)} + u_l}.
\end{aligned}
$$

From the above formula we can derive:

$$
b_l^g - b_f^g \leq b_l^g \frac{1+u_l}{\frac{1}{(1-u_m)} + u_l}.
$$

To this end, we have the following conclusions: (i) According to ***Proposition* 1**, fusing an additional opinion (e.g., M) into the original opinion (e.g., L) has great potential to boost the model's accuracy. (ii) According to ***Proposition* 2**, the above fusion may lead to accuracy deterioration, but even this is limited under mild conditions.

For the impact of intra-module design, we can utilize the same analysis approach as mentioned above, and obtain the

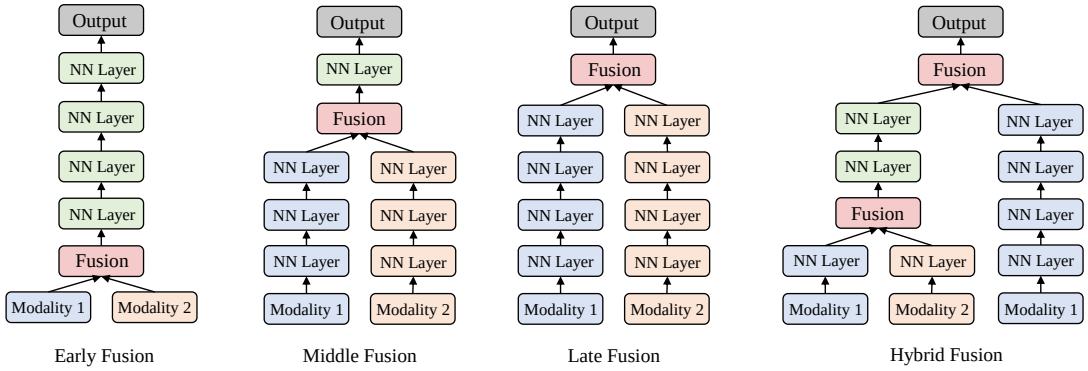

Figure 2: Examples of four different fusion types. Blue indicates extracting modality 1 features with neural networks and orange indicates modality 2. Red denotes fusing both modalities. Green represents modeling the fused feature utilizing neural networks.

Algorithm 1: Two-level Search Space Pseudo-code

```
#S1, S3, S3: List of early, middle,
    and late features for RGB and
    depth.
class Search_FuisonNet(nn.Module):
    def __init__(self, NumInputNodes,
         NumInterNodes, NumKeepEdges):
        super().__init__()
        self.fusion =FusionNet(
            NumInputNodes,
            NumInterNodes,NumKeepEdges)

    def forward(self, S1, S2, S3):
        Early_out = self.fusion(S1)
        S2.append(Early_out)
        Middle_out = self.fusion(S2)
        S3.append(Middle_out)
        Out = self.fusion(S3)
        return Out
```

Algorithm 2: The architecture search process of 3D-ADNAS.

Input: RGB and depth images (i.e. MVTec 3D-AD or Eyecandies dataset).
Output: The searched multimodal fusion architecture.
Define candidate features pool $\mathbb{F}$ and candidate fusion operations pool $\mathbb{O}$;
Initialize architecture parameters (α^{in}, α^{ex}, and β^{op}) and network parameters (w);
for $e \leftarrow 1$ **to** Epochs **do**
    Fix α^{in}, α^{ex}, and β^{op}, update w on the training set;
    Fix w, and update α^{in}, α^{ex}, and β^{op} on the validation set;
    Derive the multimodal fusion architecture based on α^{in}, α^{ex}, and β^{op};
    if achieve a higher validation accuracy. **then**
        Updating the searched multimodal fusion architecture;
    end if
end for
Return the final multimodal fusion architecture.

same conclusions as the case of the above inter-module design. For example, given the opinion of original feature (e.g., the RGB feature r): $R = \{\{b_r^n\}_{n=1}^N, u_r\}$, we aim to theoretically analyze whether the fusion of additional opinion (e.g., the depth feature d) $D = \{\{b_d^n\}_{n=1}^N, u_d\}$ will influence the model's classification accuracy. Then, following the combination rule of DST, we fuse R into D and form a new opinion $F_f = \{\{b_{f_f}^n\}_{n=1}^N, u_{f_f}\}$. Here, $b_{f_f}^n$ and u_{f_f} are new belief mass and uncertainty. After getting the opinions of the features, we can theoretically analyze the impact of intra-module fusion design on 3D-AD performance. as mentioned above of the impact of inter-module fusion design on 3D-AD performance.

B. Details of Fusion Tpyes

As shown in Fig. 2, current fusion types include early, middle, late, and hybrid fusion strategies (Xu, So, and Dai 2021). In the case of the 3D anomaly detection task, each strategy has its strengths and weaknesses. For early fusion, it is beneficial to find the fine-grained relationships between modalities by directly processing the original features. However, since the feature distributions of depth and RGB images are distinct, early fusion potentially introduces excessive noise. For middle fusion, we need to extract multi-scale features through the unimodal backbone networks, and then select part or all of the features to perform fusion. Although this fusion type retains rich multimodal information, how to seek the best depth and RGB features to perform fusion still requires a great deal of prior knowledge. For late fusion, while the independent modeling of each modality makes the model design and training comparatively easy, the representation of late-depth features is weak, and only exploiting late features to perform fusion may lead to poor model performance. The hybrid fusion type combines the strengths of early, middle, and late fusion, but it also requires a wealth of prior knowledge to design a well-crafted architecture. Therefore, designing multimodal fusion architecture with neural architecture search is a promising approach for 3D anomaly detection.

Table 1: Performance of anomaly detection evaluated by P-AUROC metric on the Eyecandies dataset. The red indicates the best results and the blue indicates the second best results.

	Method	Candy Cane	Chocolate Cookie	Chocolate Praline	Confetto	Gummy Bear	Hazelnut Truffle	Licorice Sandwish	Lollipop	Marsh-mallow	Peppermint Candy	Mean
RGB	Eyecandies$_{RGB}$	97.2	93.3	96.0	94.5	92.9	81.5	85.5	97.7	93.1	92.8	92.5
	M3DM	95.6	97.9	95.8	99.8	97.6	94.1	97.7	98.6	99.7	98.8	97.6
	3D-ADNAS	88.6	96.6	94.4	99.6	84.6	93.3	88.1	93.5	98.9	99.5	93.7
Depth	M3DM	97.7	90.3	90.2	93.0	87.5	83.2	90.9	96.8	86.8	91.8	90.8
	3D-ADNAS	97.9	89.7	73.5	76.3	77.5	55.4	68.1	63.7	73.9	90.1	76.5
RGB + Depth	Eyecandies$_{RGB-D}$	97.3	92.7	95.8	94.5	92.9	80.6	82.7	97.7	93.1	92.8	92.0
	Eyecandies$_{RGB-cD-N}$	98.0	97.9	98.2	97.8	95.1	85.3	97.1	97.8	98.5	96.7	96.2
	M3DM	97.4	98.7	96.2	99.8	96.6	94.1	97.3	98.4	99.6	98.5	97.7
	3D-ADNAS	98.5	96.7	93.5	98.9	95.9	89.9	98.6	98.6	99.3	99.7	97.0

Table 2: Performance of anomaly detection evaluated by AUPRO metric on the MVTec 3D-AD dataset. The red indicates the best results and the blue indicates the second best results.

	Method	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
Depth	Depth GAN	11.1	7.2	21.2	17.4	16.0	12.8	0.3	4.2	44.6	7.5	14.3
	Depth AE	14.7	6.9	29.3	21.7	20.7	18.1	16.4	6.6	54.5	14.2	20.3
	Depth VM	28.0	37.4	24.3	52.6	48.5	31.4	19.9	38.8	54.3	38.5	37.4
	Voxel GAN	44.0	45.3	87.5	75.5	78.2	37.8	39.2	63.9	77.5	38.9	58.3
	Voxel AE	26.0	34.1	58.1	35.1	50.2	23.4	35.1	65.8	1.5	18.5	34.8
	Voxel VM	45.3	34.3	52.1	69.7	68.0	28.4	34.9	63.4	61.6	34.6	49.2
	EasyNet	16.0	3.0	68.0	75.9	75.8	6.9	22.5	73.4	79.7	50.9	47.2
	3D-ADNAS	29.7	16.4	63.7	56.5	68.1	30.9	46.8	74.2	77.5	43.9	50.8
RGB	EasyNet	75.1	82.5	91.6	59.9	69.8	69.9	91.7	82.7	88.7	63.6	77.6
	3D-ADNAS	74.2	55.9	84.6	58.1	72.8	83.0	91.8	88.6	88.7	82.2	78.0
RGB + Depth	Depth GAN	42.1	42.2	77.8	69.6	49.4	25.2	28.5	36.2	40.2	63.1	47.4
	Depth AE	43.2	15.8	80.8	49.1	84.1	40.6	26.2	21.6	71.6	47.8	48.1
	Depth VM	38.8	32.1	19.4	57.0	40.8	28.2	24.4	34.9	26.8	33.1	33.5
	Voxel GAN	66.4	62.0	76.6	74.0	78.3	33.2	58.2	79.0	63.3	48.3	63.9
	Voxel AE	46.7	75.0	80.8	55.0	76.5	47.3	72.1	91.8	1.9	17.0	56.4
	Voxel VM	51.0	33.1	41.3	71.5	68.0	27.9	30.0	50.7	61.1	36.6	47.1
	3D-ST	95.0	48.3	98.6	92.1	90.5	63.2	94.5	98.8	97.6	54.2	83.3
	EasyNet	83.9	86.4	95.1	61.8	82.8	83.6	94.2	88.9	91.1	52.8	82.1
	3D-ADNAS	84.1	80.4	90.2	64.5	82.7	78.9	87.0	91.9	91.7	85.6	83.7

Table 3: Performance of anomaly detection evaluated by I-AUROC metric on the Eyecandies dataset.

Method	Candy Cane	Chocolate Cookie	Chocolate Praline	Confetto	Gummy Bear	Hazelnut Truffle	Licorice Sandwish	Lollipop	Marsh-mallow	Peppermint Candy	Mean
5-shot	63.6	73.4	71.6	92.0	63.5	81.2	73.8	80.9	93.4	81.5	77.5
10-shot	71.2	72.5	71.5	95.2	68.8	86.9	73.8	79.2	98.6	89.6	80.7
50-shot	65.8	97.6	90.7	97.6	80.9	53.6	91.9	93.2	100.0	97.1	86.8
Full dataset	89.6	100.0	97.0	100.0	82.7	88.2	93.1	95.0	100.0	100.0	94.6

Table 4: Performance of anomaly detection evaluated by P-AUROC metric on the Eyecandies dataset.

Method	Candy Cane	Chocolate Cookie	Chocolate Praline	Confetto	Gummy Bear	Hazelnut Truffle	Licorice Sandwish	Lollipop	Marsh-mallow	Peppermint Candy	Mean
5-shot	73.9	85.6	84.1	95.7	77.5	87.5	89.4	88.9	95.0	97.8	87.5
10-shot	75.8	72.0	82.8	98.4	77.5	85.3	91.3	97.3	91.1	97.9	86.9
50-shot	95.7	86.2	95.0	98.5	74.4	77.4	96.9	97.7	97.3	96.5	91.2
Full dataset	98.5	96.7	93.5	98.9	95.9	89.9	98.6	98.6	99.3	99.7	97.0

C. Details the connections between the nodes.

Similarly, following the approach in the **3D-ADNAS Method Section**, we can also obtain the optimal connections (identified by α^{in}) between input nodes and intermediate nodes in each cell. For the k-th intermediate node, we need to select

Table 5: Performance of anomaly detection evaluated by AUPRO metric on the Eyecandies dataset.

Method	Candy Cane	Chocolate Cookie	Chocolate Praline	Confetto	Gummy Bear	Hazelnut Truffle	Licorice Sandwish	Lollipop	Marsh-mallow	Peppermint Candy	Mean
5-shot	76.3	58.4	56.7	86.0	53.3	65.3	67.8	64.2	85.3	91.1	70.4
10-shot	78.5	63.7	64.3	93.9	66.3	65.9	69.9	86.7	86.5	91.2	76.7
50-shot	84.0	75.6	78.8	94.4	70.6	68.5	88.7	87.6	91.5	94.9	83.5
Full dataset	94.5	89.1	82.7	95.8	85.7	74.8	91.1	90.7	96.4	97.2	89.8

Table 6: Performance of anomaly detection evaluated by I-AUROC metric on the MVTec 3D-AD dataset.

Method	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
5-shot	93.0	89.3	76.9	82.5	73.7	84.2	80.6	69.6	97.7	78.1	82.6
10-shot	92.7	92.8	79.8	86.1	78.6	86.1	77.7	78.8	98.9	76.3	84.8
50-shot	97.6	100.0	85.5	95.9	85.1	89.7	79.8	79.5	98.6	78.4	89.0
Full dataset	99.7	100.0	97.1	98.6	96.6	94.8	89.7	87.3	100.0	86.7	95.1

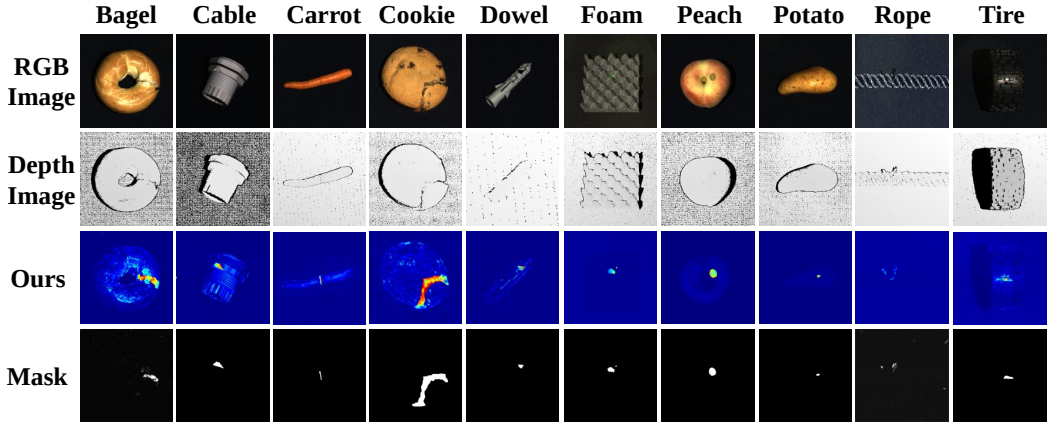

Figure 3: Visualization results of 3D-ADNAS on the MVTec 3D-AD dataset.

two features from the candidate pool feature ($\mathbb{F}$) which consists of the inputs of the cell and the predecessor nodes output of the current node as the inputs of the k-th intermediate node according to architectural parameters α^{in}. That is, it is required to optimize α^{in} and select the features with largest α^{in} values as optimal solutions. To solve this problem, we utilize the continuous relaxation strategy of DARTS to convert the discrete feature selection problem into the continuous search problem, and then use gradient method to solve it. To achieve this, we reformulate the intermediate node inputs through weighted summation of all candidate multimodal features in $\mathbb{F}$ as:

$$\tilde{Y}_i = \sum_{f_s \in \mathbb{F}} \frac{exp(\alpha_i^{in_i})}{\sum_{j=1}^{\mathbb{F}} exp(\alpha_i^{in_j})} \cdot f_s, \tag{6}$$

where $\tilde{Y}_i$ denotes the input features of each intermediate node, $i \in [1, 2]$, and f_s denotes the feature in $\mathbb{F}$. Accordingly, the two features serving as inputs of each intermediate node can be determined by:

$$(f_s^j, f_s^h) = \underset{i \in [1, \mathbb{F}]}{\arg\max} \, (\alpha_1^{in_i}, \alpha_2^{in_i}), \tag{7}$$

where (f_s^j, f_s^h) denotes the choice of the j-th and h-th features from $\mathbb{F}$ as the intermediate node inputs.

D. P-AUROC Scores on Eyecandies

Table 1 shows the comparison between 3D-ADNAS and the SOTA methods on the Eyecandies dataset (Bonfiglioli et al. 2022; Wang et al. 2023). It is obvious that 3D-ADNAS still exhibits an excellent performance. In particular, 3D-ADNAS achieves a 97% P-AUROC score when training with RGB and 3D depth images. These results can demonstrate the key role of the multimodal fusion architecture design in 3D-AD. This means that the advancement of 3D-AD through improved multimodal fusion architecture design is a feasible method.

E. AUPRO Scores on MVTec 3D-AD

Table 2 reports the AUPRO scores of 3D-ADNAS on the MVTec 3D-AD dataset (Bergmann et al. 2022). It is obvious that our method consistently outperforms the baseline models when training with RGB and depth images, such as 3D-ADNAS achieves 0.4% and 1.6% higher than 3D-ST (Bergmann and Sattlegger 2023) and EasyNet (Chen et al. 2023) in terms of AUPRO scores, respectively. This result

suggests that 3D-ADNAS can find a helpful multimodal fusion architecture to foster 3D-AD performance. In summary, the design of multimodal fusion architecture/topology indeed plays an important role in the 3D-AD development.

F. Detailed Results of Few-shot Setting

In the **Evaluating the Improved Fusion Topology Section** of the main paper, we obtain the experimental results of 3D-ADNAS on the Eyecandies and MVTec datasets for the few-shot setting (Duan et al. 2023; Kim et al. 2024), and the detailed results for all categories are shown in Tables 3 to 6. Specifically, we randomly choose 5, 10, and 50 images (RGB and Depth images) as training data from each category, and evaluate the 3D-ADNAS model performance on the full testing dataset. From these tables, we can find that our method exhibits a strong potential in the few-shot setting. For example, on the 50-shot setting, 3D-ADNAS achieves 89.0% on MVTec 3D-AD and 86.8% on Eyecandies in terms of I-AUROC scores; on the 10-shot setting, 3D-ADNAS achieves 84.8% on MVTec 3D-AD and 80.7% on Eyecandies in terms of I-AUROC scores. This implies that a friendly multimodal fusion topology is still effective for the few-shot 3D-AD task. In the future, we can promote the development of the few-shot 3D-AD task from the perspective of improving the few-shot 3D-AD architecture in a similar way.

G. Visualization Results

Fig. 3 shows the visualization results of 3D-ADNAS on the MVTec 3D-AD dataset, which demonstrates that optimizing the multimodal fusion architecture design can improve 3D-AD performance. From Fig. 3, we can observe that 3D-ADNAS can correctly locate the abnormal regions by exploiting the color features of RGB as well as the geometric features of the depth image. Although some of the RGB and depth images involve mutually interfering features, our proposed 3D-ADNAS method can automatically seek the optimal features to perform fusion from numerous interfering features. This indicates that the improvement of 3D-AD can be boosted not only by the learning paradigm but also through improved multimodal fusion architecture design.

References

Bergmann, P.; Jin, X.; Sattlegger, D.; and Steger, C. 2022. The MVTec 3D-AD Dataset for Unsupervised 3D Anomaly Detection and Localization. In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 202–213.

Bergmann, P.; and Sattlegger, D. 2023. Anomaly detection in 3d point clouds using deep geometric descriptors. In *WACV*, 2613–2623.

Bishop, C. M.; and Nasrabadi, N. M. 2006. *Pattern recognition and machine learning*, volume 4. Springer.

Bonfiglioli, L.; Toschi, M.; Silvestri, D.; Fioraio, N.; and De Gregorio, D. 2022. The eyecandies dataset for unsupervised multimodal anomaly detection and localization. In *Proceedings of the Asian Conference on Computer Vision*, 3586–3602.

Chen, R.; Xie, G.; Liu, J.; Wang, J.; Luo, Z.; Wang, J.; and Zheng, F. 2023. Easynet: An easy network for 3d industrial anomaly detection. In *ACM MM*, 7038–7046.

Duan, Y.; Hong, Y.; Niu, L.; and Zhang, L. 2023. Few-shot defect image generation via defect-aware feature manipulation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 571–578.

Han, Z.; Zhang, C.; Fu, H.; and Zhou, J. T. 2021. Trusted Multi-View Classification. In *International Conference on Learning Representations*.

Han, Z.; Zhang, C.; Fu, H.; and Zhou, J. T. 2022. Trusted multi-view classification with dynamic evidential fusion. *IEEE transactions on pattern analysis and machine intelligence*, 45(2): 2551–2566.

Kim, S.; An, S.; Chikontwe, P.; Kang, M.; Adeli, E.; Pohl, K. M.; and Park, S. H. 2024. Few Shot Part Segmentation Reveals Compositional Logic for Industrial Anomaly Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 8591–8599.

Kingma, D. P.; and Welling, M. 2014. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations*.

Liu, Y.-T.; Pal, N. R.; Marathe, A. R.; and Lin, C.-T. 2017. Weighted fuzzy Dempster–Shafer framework for multimodal information integration. *IEEE Transactions on Fuzzy Systems*, 26(1): 338–352.

Wang, Y.; Peng, J.; Zhang, J.; Yi, R.; Wang, Y.; and Wang, C. 2023. Multimodal industrial anomaly detection via hybrid fusion. In *CVPR*, 8032–8041.

Xu, C.; Zhao, W.; Zhao, J.; Guan, Z.; Song, X.; and Li, J. 2022. Uncertainty-aware multiview deep learning for internet of things applications. *IEEE Transactions on Industrial Informatics*, 19(2): 1456–1466.

Xu, Z.; So, D. R.; and Dai, A. M. 2021. Mufasa: Multimodal fusion architecture search for electronic health records. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 10532–10540.