# Confidence-based Visual Dispersal for Few-Shot Unsupervised Domain Adaptation


北京信息科学与技术国家研究中心
BEIJING NATIONAL RESEARCH CENTER FOR INFORMATION SCIENCE AND TECHNOLOGY

Yizhe Xiong[1,2,3*], Hui Chen[2,†], Zijia Lin[1], Sicheng Zhao[2], Guiguang Ding[1,2,†]

[1]Tsinghua University [2]BNRist [3]Hangzhou Zhuoxi Institute of Brain and Intelligence

涿溪实验室 ZHUO XI LABORATORY


ICCV23 PARIS

## Introduction & Motivation

In real-world scenarios, providing abundant labels even for the source domain can be difficult and costly, leading to the Few-shot Unsupervised Domain Adaptation (FUDA) problem, where source samples are scarcely labeled. Existing methods overlook that the sparse label setting hinders learning reliable knowledge for transfer, resulting in performance drops.

We present the Confidence-based Visual Dispersal Transfer learning method (C-VisDiT) for FUDA, aiming to comprehensively consider the importance of each sample during transfer based on its confidence.

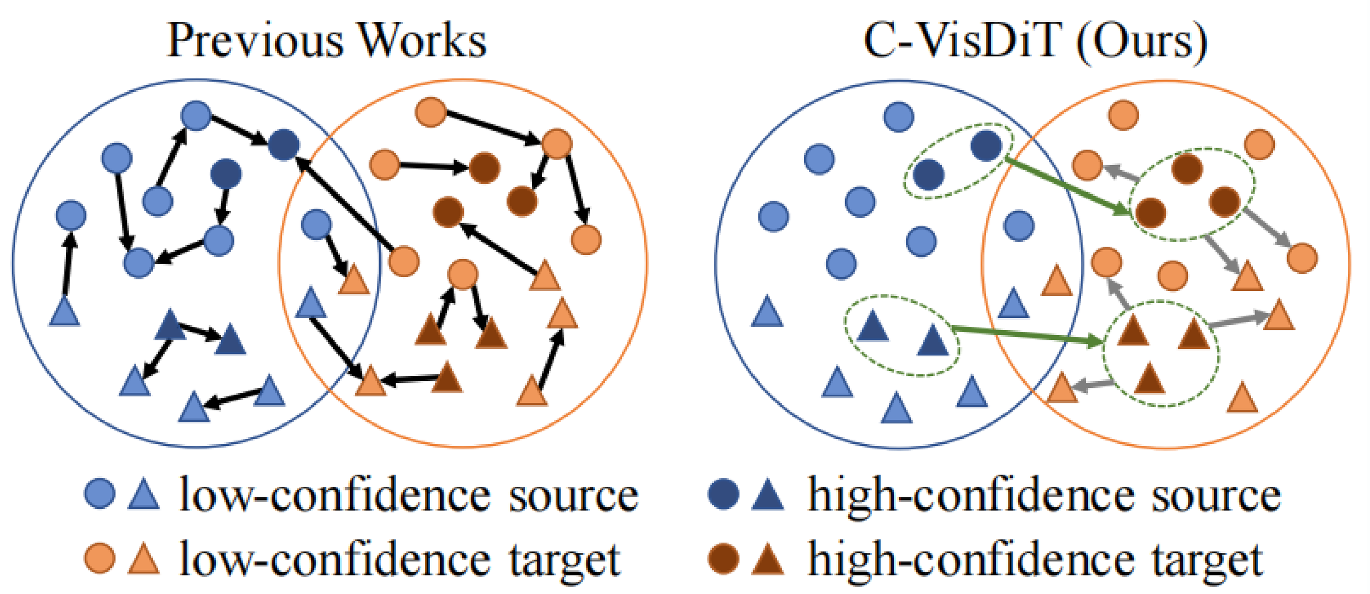



Figure 1: We address the task of FUDA. Left: Existing FUDA methods implicitly align source and target domains without considering the confidence of samples. Right: Our C-VisDiT transfers knowledge from high-confidence source samples and high-confidence target samples to enhance model adaptation.

## Methodology

### ✓Cross-domain Visual Dispersal (X-VD)

We aim to transfer high-confidence source knowledge by pulling target samples towards corresponding high-confidence source samples with similar semantics.

- ➢ Nearest Source Matching: $M = \{m_{ij}\}_{N_{ut}\times N_{ls}} \quad m_{ij} = \|F(\mathbf{x}_i^{ut}) - F(\mathbf{x}_j^{ls})\|_2$
- ➢ High-confidence Target Sampling: $O_{conf} = \arg\mathrm{TopK}(-\mathcal{H}(\mathbf{x}_i^{ut}), N_{\mathrm{X}})$
- ➢ Cross-domain Knowledge Transferring via Visual Dispersal:

$$\mathbf{x}_i^{\mathrm{X}} = \beta\bar{\mathbf{x}}_i^{ut} + (1-\beta)\bar{\mathbf{x}}_i^{ut\leftarrow ls} \quad y_i^{\mathrm{X}} = \beta y_i^{ut} + (1-\beta)y_i^{ut\leftarrow ls}$$

### ✓Intra-domain Visual Dispersal (I-VD)

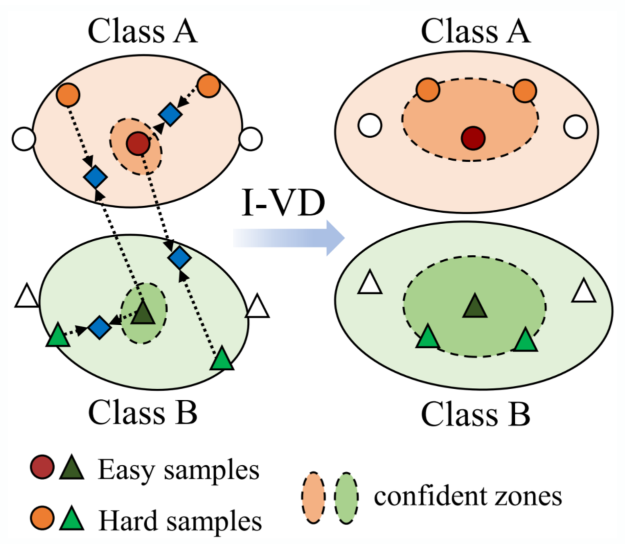
We aim to enhance the classification of difficult target samples with the guidance of "easy" target samples.

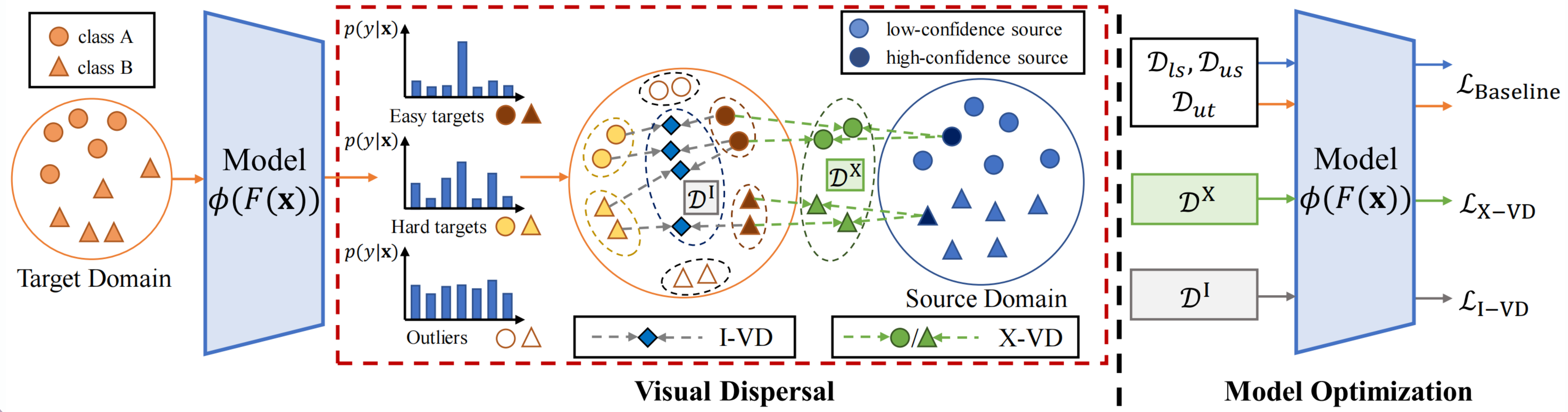
- ➢ Confidence-based Target Domain Splitting:

$$O_1 = O[0{:}N_1],\ O_2 = O[N_1{:}N'],\ O_3 = O[N'{:}N_{ut}]$$

- ➢ Intra-domain Sample Guidance via Visual Dispersal:

$$\mathbf{x}_i^{\mathrm{I}} = \beta\bar{\mathbf{x}}_j^{E} + (1-\beta)\bar{\mathbf{x}}_i^{H} \quad y_i^{\mathrm{I}} = \beta y_j^{E} + (1-\beta)y_i^{H}$$




## Visualization Results

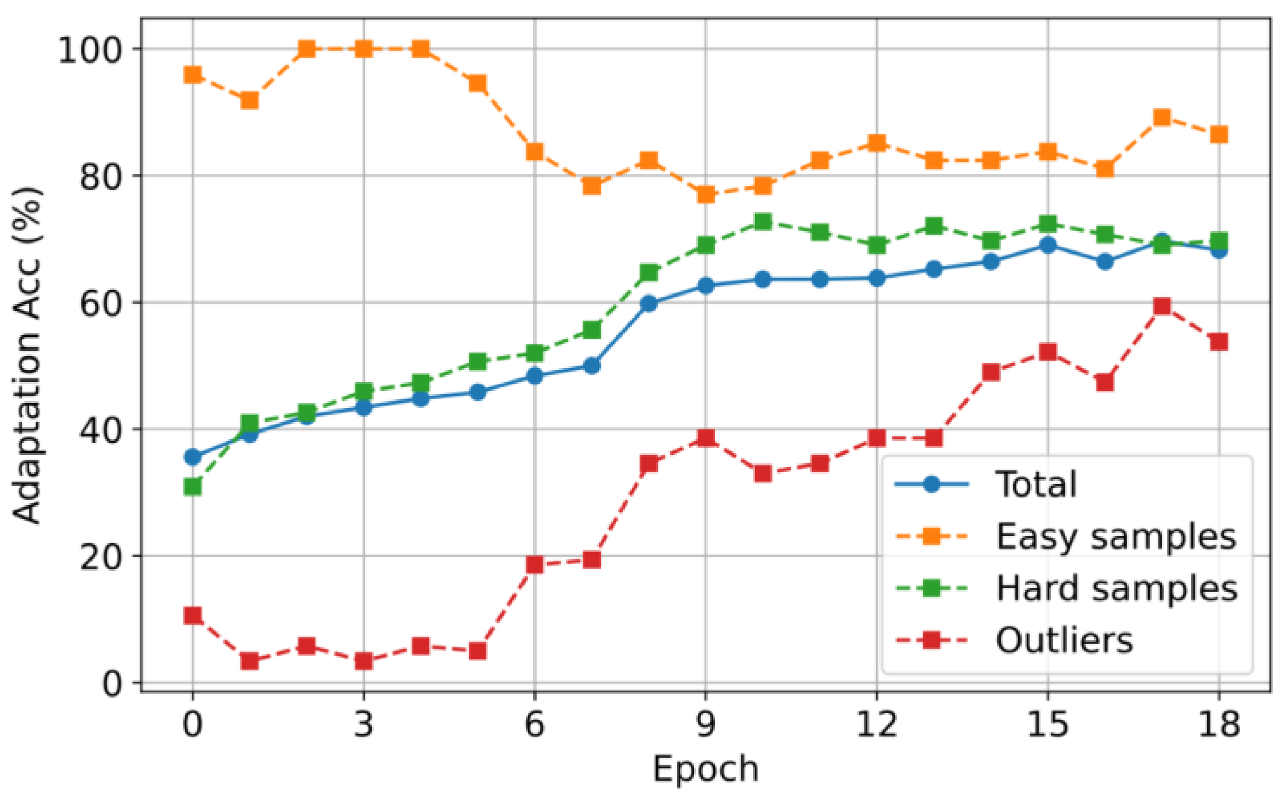



Figure 4: Accuracy change of target samples with different learning difficulties.

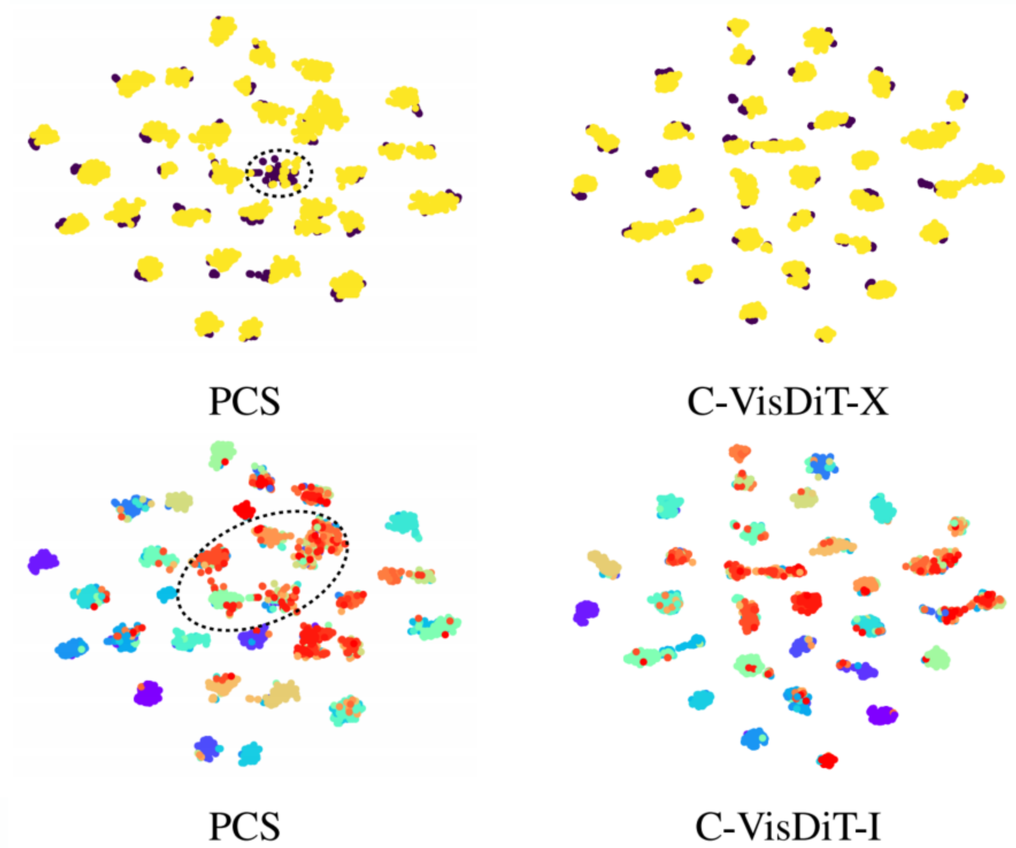
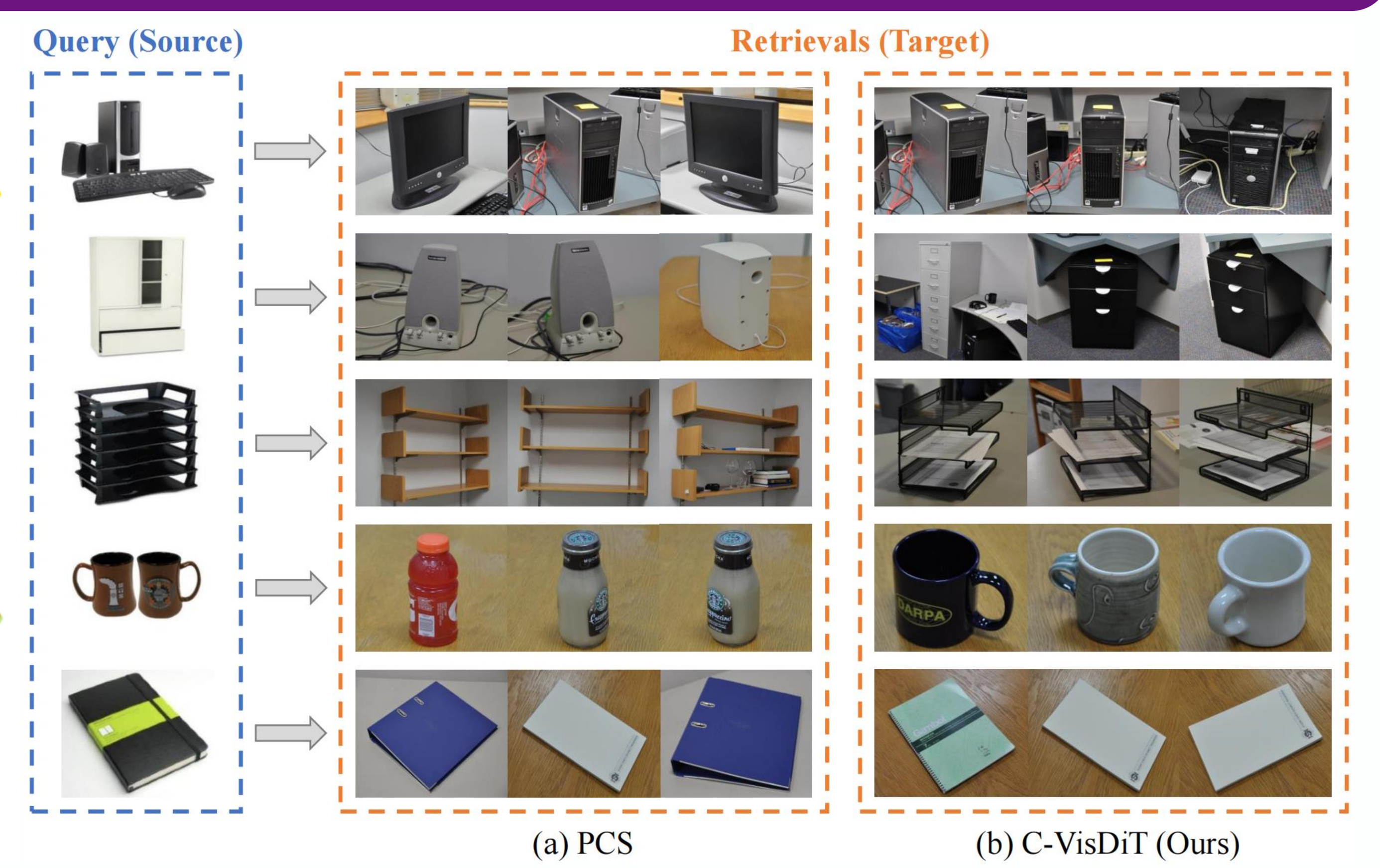



Figure 5: Office-31 t-SNE visualization results.




(Appendix) Figure 4: Image retrieval examples.

## Experimental Results

Table 1: Adaptation accuracy (%) comparison on 1-shot / 3-shots labeled source per class on the Office-31 dataset.

| Method | A→D | A→W | D→A | D→W | W→A | W→D | Avg |
|---|---|---|---|---|---|---|---|
| Source Only | 31.3 / 49.0 | 19.6 / 43.3 | 41.3 / 55.7 | 58.7 / 81.8 | 39.9 / 49.8 | 61.9 / 81.7 | 42.1 / 60.2 |
| MME [34] | 21.5 / 51.0 | 12.2 / 54.6 | 23.1 / 60.2 | 60.9 / 89.7 | 14.0 / 52.3 | 62.4 / 91.4 | 32.3 / 66.5 |
| CDAN [25] | 11.2 / 43.7 | 6.2 / 50.1 | 9.1 / 65.1 | 54.8 / 91.6 | 10.4 / 57.0 | 41.6 / 89.8 | 22.2 / 66.2 |
| MDDIA [16] | 45.0 / 62.9 | 54.5 / 65.4 | 55.6 / 67.9 | 84.4 / 93.3 | 53.4 / 70.3 | 79.5 / 93.2 | 62.1 / 75.5 |
| CDS [17] | 52.6 / 65.1 | 55.2 / 68.8 | 65.7 / 71.2 | 76.6 / 88.1 | 59.7 / 71.0 | 73.3 / 87.3 | 63.9 / 75.3 |
| PCS [50] | 60.2 / 78.2 | 69.8 / 82.9 | **76.1** / 76.4 | 90.6 / 94.1 | 71.2 / 76.3 | 91.8 / 96.0 | 76.6 / 84.0 |
| C-VisDiT (Ours) | **74.1 / 82.7** | **72.3 / 86.0** | 75.7 / **76.5** | **93.2 / 95.0** | **76.4 / 76.9** | **94.2 / 97.0** | **81.0 / 85.7** |

Table 2: Adaptation accuracy (%) comparison on 3% and 6% labeled source samples per class on the Office-Home dataset.

| Method | Ar →Cl | Ar →Pr | Ar →Rw | Cl →Ar | Cl →Pr | Cl →Rw | Pr →Ar | Pr →Cl | Pr →Rw | Rw →Ar | Rw →Cl | Rw →Pr | Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **3% labeled source** | | | | | | | | | | | | | |
| Source Only | 22.5 | 36.5 | 41.1 | 18.5 | 29.7 | 28.6 | 27.2 | 25.9 | 38.4 | 33.5 | 20.3 | 41.4 | 30.3 |
| MME [34] | 4.5 | 15.4 | 25.0 | 28.7 | 34.1 | 37.0 | 25.6 | 25.4 | 44.9 | 39.3 | 29.0 | 52.0 | 30.1 |
| CDAN [25] | 5.0 | 8.4 | 11.8 | 20.6 | 26.1 | 27.5 | 26.6 | 27.0 | 40.3 | 38.7 | 25.5 | 44.9 | 25.2 |
| MDDIA [16] | 21.7 | 37.3 | 42.8 | 29.4 | 43.9 | 44.2 | 37.7 | 29.5 | 51.0 | 47.1 | 29.2 | 56.4 | 39.1 |
| CDS [17] | 43.8 | 55.5 | 60.2 | 51.5 | 56.4 | 59.6 | 51.3 | 46.4 | 64.5 | 62.2 | 52.4 | 70.2 | 56.2 |
| PCS [50] | 42.1 | 61.5 | 63.9 | 52.3 | 61.5 | 61.4 | 58.0 | 47.6 | 73.9 | 66.0 | **52.5** | 75.6 | 59.7 |
| C-VisDiT (Ours) | **44.1** | **66.8** | **67.0** | **54.9** | **66.4** | **66.8** | **60.5** | **47.9** | **75.7** | **67.2** | 51.6 | **78.8** | **62.3** |
| **6% labeled source** | | | | | | | | | | | | | |
| Source Only | 26.5 | 41.3 | 46.7 | 29.3 | 40.4 | 37.9 | 35.5 | 31.6 | 57.2 | 46.2 | 32.7 | 59.2 | 40.4 |
| MME [34] | 27.6 | 43.2 | 49.5 | 41.1 | 46.6 | 49.5 | 43.7 | 30.5 | 61.3 | 54.9 | 37.3 | 66.8 | 46.0 |
| CDAN [25] | 26.2 | 33.7 | 44.5 | 34.8 | 42.9 | 44.7 | 42.9 | 36.0 | 59.3 | 54.9 | 40.1 | 63.6 | 43.6 |
| MDDIA [16] | 25.1 | 44.5 | 51.9 | 35.6 | 46.7 | 50.3 | 48.3 | 37.1 | 64.5 | 58.2 | 36.9 | 68.4 | 50.3 |
| CDS [17] | 45.4 | 60.4 | 65.5 | 54.9 | 59.2 | 63.8 | 55.4 | **49.0** | 71.6 | 66.6 | **54.1** | 75.4 | 60.1 |
| PCS [50] | 46.1 | 65.7 | 69.2 | 57.1 | 64.7 | 66.2 | 61.4 | 47.9 | 75.2 | 67.0 | 53.9 | 76.6 | 62.6 |
| C-VisDiT (Ours) | **46.5** | **69.8** | **72.5** | **60.2** | **71.2** | **71.2** | **64.1** | **49.0** | **78.5** | **69.1** | 52.8 | **80.1** | **65.4** |

| Method | SO | MME [34] | CDAN [25] | CDS [17] | PCS [50] | C-VisDiT |
|---|---|---|---|---|---|---|
| Acc. | 42.6 | 69.4 | 61.5 | 69.4 | 79.0 | **80.5** |

Table 3: Adaptation accuracy (%) comparison on 1% labeled source samples per class on the VisDA-C dataset.

| Method | SO | MME [34] | CDAN [25] | CDS [17] |
|---|---|---|---|---|
| 1-shot | 13.1 | 17.5 | 14.6 | 21.5 |
| 3-shots | 27.1 | 29.1 | 27.3 | 42.5 |
| **Method** | **PCS [50]** | **C-VisDiT** | **BrAD [12]** | **BrAD+C-VisDiT** |
| 1-shot | 34.7 | **35.9** | 49.7 | **51.0** |
| 3-shots | 46.1 | **48.1** | 60.9 | **63.4** |

Table 4: Adaptation accuracy (%) comparison on 1-shot and 3-shots labeled source per class on the DomainNet dataset.

( In Table 3 and 4, "SO" denotes training with only labeled source samples, or Source Only)

## Conclusion

- We propose a Confidence-based Visual Dispersal Transfer learning method (C-VisDiT) for FUDA, which simultaneously takes the reliability of the source samples and the learning difficulty of target samples into account.
- We introduce X-VD to avoid the negative impact of source knowledge with low confidence. We employ I-VD to boost the learning of hard target samples with the guidance of easy ones.
- Experimental results on four popular benchmark datasets show that our C-VisDiT establishes new state-of-the-art results for all datasets.