# Jiaxin Chen

## Personal Information

| | |
|---|---|
| **Status:** | **Post-doctor** |
| **Program:** | **Computer Science and Engineering** |
| **School:** | **Tandon School of Engineering, New York University** |
| **Period:** | **From 2017-07 to 2018-07** |

## Biography

**I was a post-doctor at New York University and advised by Professor Yi Fang. During my post doctoral period, I was a research assistant in NYU Multimedia and Visual Computing (MMVC) Lab. I am broadly interested in 3D Computer Vision and Deep Learning. Now I am a Research Scientist at Inception Institute of Artificial Intelligence, Abu Dhabi.**

**Research Project:** Deep Cross-modality Adaptation via Semantics Preserving Adversarial Learning for Sketch-based 3D Shape Retrieval

## Description

Due to the large cross-modality discrepancy between 2D sketches and 3D shapes, retrieving 3D shapes by sketches is a significantly challenging task. To address this problem, we propose a novel framework to learn a discriminative deep cross-modality adaptation model. We first separately adopt two metric networks to learn modality-specific discriminative features based on an importance-aware metric learning method. Then, we explicitly introduce a cross-modality transformation network to compensate for the divergence between two modalities. We develop an adversarial learning based method to train the transformation model. Experimental results on the SHREC 2013 and SHREC 2014 datasets clearly show the superior retrieval performance of our proposed model, compared to the state-of-the-art approaches.

## Method

We propose a novel model, namely Deep Cross-modality Adaptation (DCA), for sketch-based 3D shape retrieval. We first construct two separate deep convolutional neural networks (CNNs) and metric networks, one for sketches and the other for 3D shapes, to learn discriminative modality- specific features for each modality via importance-aware metric learning (IAML). In order to reduce the large cross-modality divergence between learned features of sketches and 3D shapes, we explicitly introduce a cross-modality transformation network, to transfer features of sketches into the feature space of 3D shapes. An adversarial learning method with class-aware cross-modality mean discrepancy minimization (CMDM-AL) is developed to train the transformation network. IAML is also applied to the transformed data, in order to further preserve semantic structures of sketch data after adaptation.

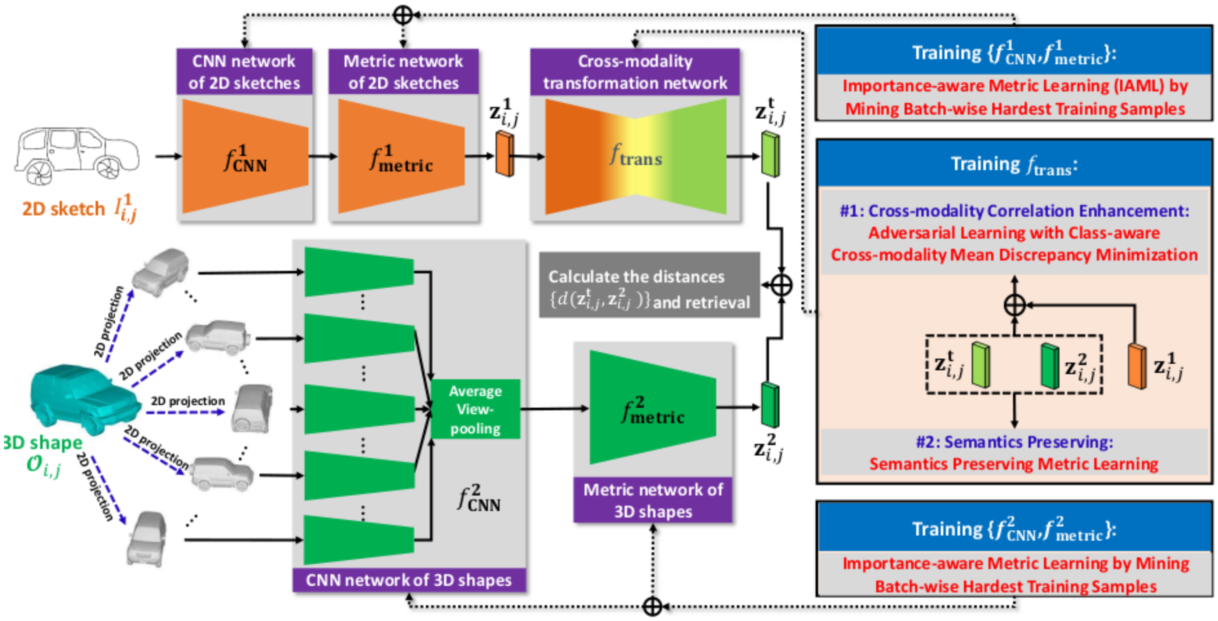


Figure 1: Framework of our proposed method. Our model consists of the CNN network and metric network of 2D sketches, the CNN network and metric network metric of rendered images of 3D shapes, together with the cross-modality transformation network. The CNN and metric networks for each single modality (i.e., 2D sketches or 3D shapes) is trained by importance-aware metric learning through mining the hardest training samples. The cross transformation network is trained by enforcing features of sketches to be semantics preserving after adaptation. Simultaneously, an adversarial learning with cross-modality mean discrepancy minimization is employed to enhance both the local and holistic correlations between data distributions of transformed features of sketches and features of 3D shapes.

# Results

In order to evaluate the performance of our method, we conduct experiments on two widely used benchmark datasets for sketch-based 3D shape retrieval: i.e., SHREC 2013 and SHREC 2014. We report the precision-recall curve. Extensive experimental results on two benchmark datasets demonstrated the superiority of the propose method, compared to the state-of-the-art approaches.

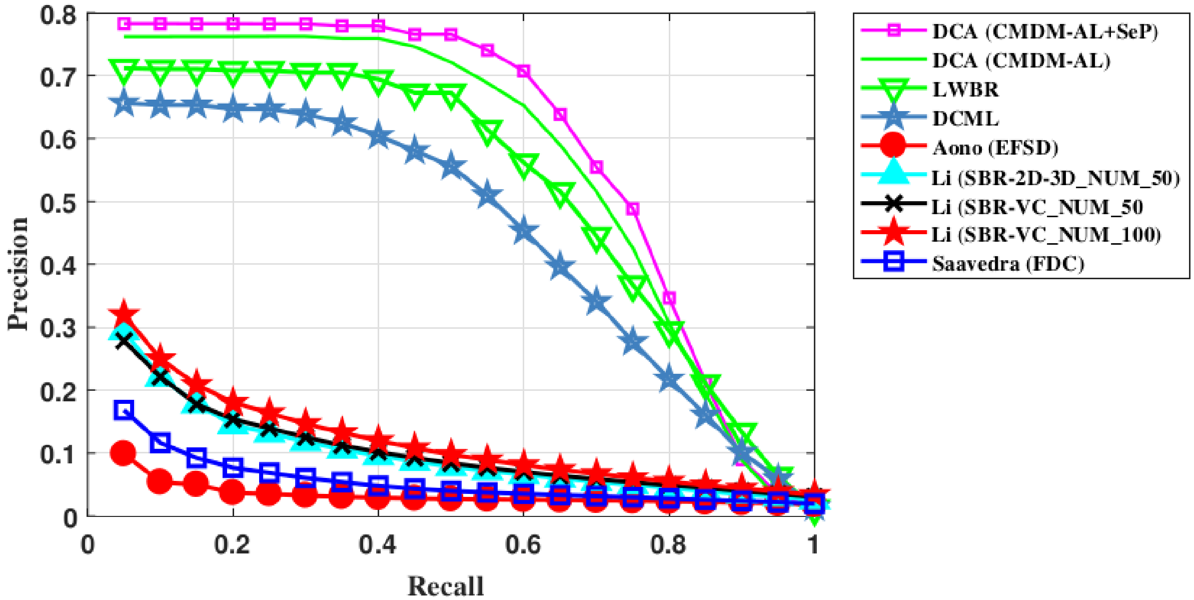


Figure 2: The precision-recall curves of various methods on SHREC 2013.

| Methods | NN | FT | ST | E | DCG | mAP |
|---|---|---|---|---|---|---|
| CDMR [4] | 0.279 | 0.203 | 0.296 | 0.166 | 0.458 | 0.250 |
| SBR-VC [12] | 0.164 | 0.097 | 0.149 | 0.085 | 0.348 | 0.114 |
| SP [23] | 0.017 | 0.016 | 0.031 | 0.018 | 0.240 | 0.026 |
| FDC [12] | 0.110 | 0.069 | 0.107 | 0.061 | 0.307 | 0.086 |
| Siamese [28] | 0.405 | 0.403 | 0.548 | 0.287 | 0.607 | 0.469 |
| CAT-DTW [32] | 0.235 | 0.135 | 0.198 | 0.109 | 0.392 | 0.141 |
| KECNN [25] | 0.320 | 0.319 | 0.397 | 0.236 | 0.489 | - |
| DCML [2] | 0.650 | 0.634 | 0.719 | 0.348 | 0.766 | 0.674 |
| LWBR [30] | 0.712 | 0.725 | 0.785 | 0.369 | 0.814 | 0.752 |
| **DCA (SeP)** | 0.009 | 0.015 | 0.027 | 0.014 | 0.231 | 0.034 |
| **DCA (CMDM-AL)** | 0.762 | 0.776 | 0.812 | 0.370 | 0.842 | 0.795 |
| **DCA (CMDM-AL+SeP)** | **0.783** | **0.796** | **0.829** | **0.376** | **0.856** | **0.813** |

Figure 3: Performance on SHREC 2013, compared with the state-of-the-art methods.

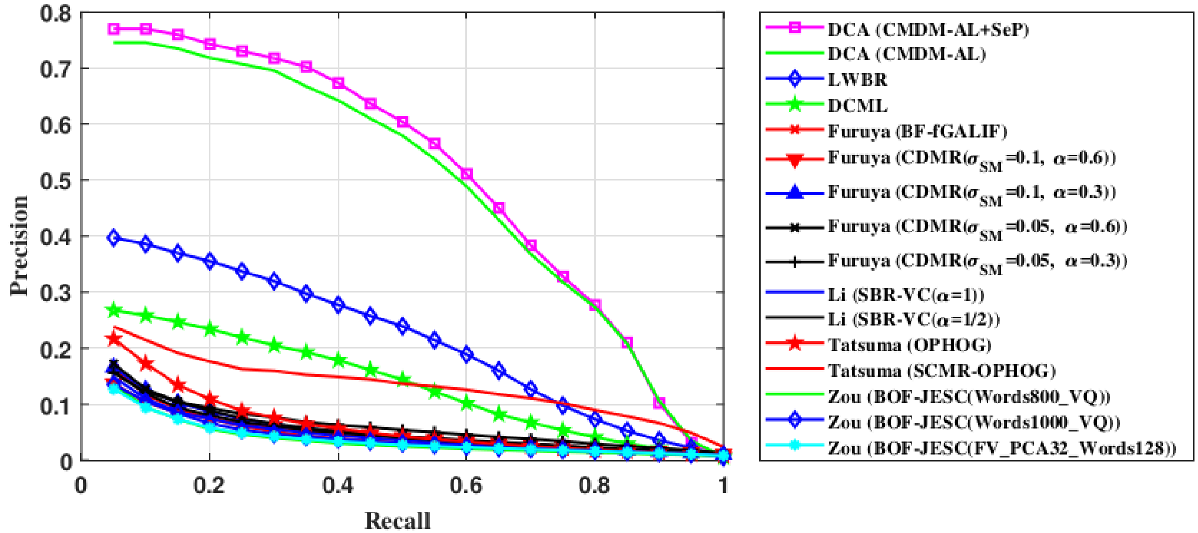


Figure 4: The precision-recall curves of various methods on SHREC 2014.

| **Methods** | **NN** | **FT** | **ST** | **E** | **DCG** | **mAP** |
|---|---|---|---|---|---|---|
| CDMR [4] | 0.109 | 0.057 | 0.089 | 0.041 | 0.328 | 0.054 |
| SBR-VC [12] | 0.095 | 0.050 | 0.081 | 0.037 | 0.319 | 0.050 |
| DB-VLAT [26] | 0.160 | 0.115 | 0.170 | 0.079 | 0.376 | 0.131 |
| CAT-DTW [32] | 0.137 | 0.068 | 0.102 | 0.050 | 0.338 | 0.060 |
| Siamese [28] | 0.239 | 0.212 | 0.316 | 0.140 | 0.496 | 0.228 |
| DCML [2] | 0.272 | 0.275 | 0.345 | 0.171 | 0.498 | 0.286 |
| LWBR [30] | 0.403 | 0.378 | 0.455 | 0.236 | 0.581 | 0.401 |
| AlexNet-DCA | 0.498 | 0.464 | 0.513 | 0.294 | 0.627 | 0.502 |
| VGG-16-DCA | 0.682 | 0.698 | 0.723 | 0.375 | 0.783 | 0.711 |
| **DCA (SeP)** | 0.018 | 0.020 | 0.028 | 0.007 | 0.266 | 0.030 |
| **DCA (CMDM-AL)** | 0.745 | 0.766 | 0.808 | 0.392 | 0.845 | 0.782 |
| **DCA (CMDM-AL+SeP)** | **0.770** | **0.789** | **0.823** | **0.398** | **0.859** | **0.803** |

Figure 5: Performance on SHREC 2014, compared with the state-of-the-art methods.