Cross-Dataset Adaptation for Visual Question Answering

Wei-Lun (Harry) Chao*, Hexiang Hu*, and Fei Sha U. of Southern California

This work is partially supported by USC Graduate Fellowship, NSF IIS-1065243, 1451412, 1513966/1632803, 1208500, CCF-1139148, a Google Research Award, an Alfred. P. Sloan Research Fellowship and ARO# W911NF-12-1-0241 and W911NF-15-1-0484.

Highlights

- ***Analyze the bias in Visual QA datasets that hinders knowledge transfer***
- ***Propose a domain adaptation algorithm:***
  (1) transform target data to match distribution
  (2) leverage source domain's Visual QA knowledge
- ***Evaluate across 5 popular Visual QA datasets (with no need to re-train source domains' models)***

Dataset bias

- ***Visual7W [CVPR 2016] vs. VQA [ICCV 2015]***

Question: Who leads the parade?

Candidates: The mayor. The governor. The clowns. Motorcycle cop.

Question: What type of bike is this?

Candidates: No. Bike for two. Kingfish. Motorcycle.

- ***Name that dataset!***

Image (I)	Question (Q)	Target A (T)	Decoy A (D)	Q + T + D
52.3	76.3	74.7	95.8	**97.5**

- ***Poor cross-dataset generalization*** (MLP model [6])

Training \ Testing	Visual7W	VQA-
Visual7W	**65.7**	28.1
VQA-	53.4	**55.6**

VQA-: remove Yes/No examples

Methodology

Training (Source domain: SD)

Q: Where is the picture taken?

Visual QA system → **T:** Baseball field

Adaptation

Testing (Target domain: TD)

Q: Where are they? **T:** Baseball court

$\left\{\left(\mathrm{I}^{(n)},\mathrm{Q}^{(n)},\mathbf{A}^{(n)}\right)\right\}_{n=1}^{N} \Rightarrow f_{\mathrm{SD}}(\mathrm{I},\mathrm{Q},\mathrm{C}) \in \mathrm{R}$

$\mathbf{A} = \{\mathrm{T},\mathrm{D}_1,\cdots,\mathrm{D}_K\}$

Q: What is the animal?

$\begin{cases} \mathrm{C}=\mathrm{T},\ f_{\mathrm{SD}}(\mathrm{I},\mathrm{Q},\mathrm{C}) \uparrow & \text{Horse} \\ \mathrm{C}\neq\mathrm{T},\ f_{\mathrm{SD}}(\mathrm{I},\mathrm{Q},\mathrm{C}) \downarrow & \text{Dog} \end{cases}$

$\mathrm{argmax}_{\mathrm{C}\in\mathbf{A}}\ f_{\mathrm{SD}}(\mathrm{I},\mathrm{Q},\mathrm{C})$

Q & C (a candidate) are encoded by BoW or RNN.

The source's model won't be re-trained!

Learn m_Q, m_C from TD to SD

$\left\{\left(\mathrm{I}^{(n)},\mathrm{Q}^{(n)},\mathbf{A}^{(n)}\right)\right\}_{n=1}^{N} \quad f_{\mathrm{SD}}(\mathrm{I},\mathrm{Q},\mathrm{C})$

$P_{\mathrm{SD}}(\mathrm{Q},\mathrm{T}) \Downarrow \quad f_{\mathrm{SD}}\left(\mathrm{I}, m_Q(\mathrm{Q}), m_C(\mathrm{C})\right)$

$\approx f_{\mathrm{SD}}\left(m_Q(\mathrm{Q}), m_C(\mathrm{C})\right)$

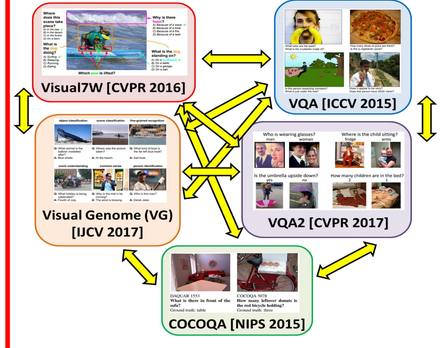
$P_{\mathrm{TD}}\left(m_Q(\mathrm{Q}), m_C(\mathrm{T})\right) \Uparrow \quad \mathrm{C}=\mathrm{T},\uparrow \quad \mathrm{C}\neq\mathrm{T},\downarrow$

- minimize mismatch
- maximize performance

$\left\{\left(\mathrm{Q}^{(m)},\mathbf{A}^{(m)}\right)\right\}_{m=1}^{M}$

Experiments

5 Visual QA datasets [4,16,27,33,50] (all use MSCOCO images)



VQA-, VQA2-: remove Yes/No examples

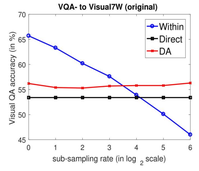
VQA- [4] → Visual7W [50] (original datasets)

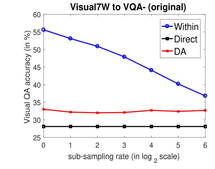
Direct	CORAL [36]	ADDA [41]	[T+D]	[Q+T]	[Q+T+D]	Within
53.4	53.4	54.1	55.7	55.2	**58.5**	65.7

Visual7W [50] → VQA- [4] (original datasets)

28.1	26.9	29.2	33.6	29.4	**35.2**	55.6

Robust to TD size

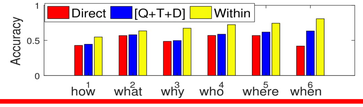




Benefit all Q types

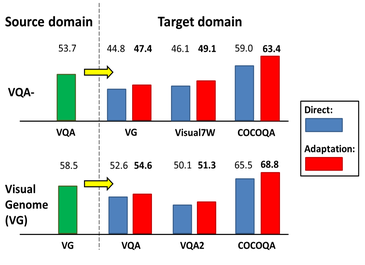
(VQA- → Visual7W)

"When" raises >20%



Transfer across 5 "revised" datasets [6]: TD with [Q+T+D] and 1/16 data

	Visual7W			VQA⁻			VG			COCOQA			VQA2⁻		
Training/Testing	Direct	DA	Within	Direct	DA	Within	Direct	DA	Within	Direct	DA	Within	Direct	DA	Within
Visual7W	52.0	-	-	45.6	48.0	43.1	49.1	49.4	48.0	58.0	63.1	65.2	43.9	45.5	43.6
VQA⁻	46.1	49.1	39.7	53.7	-	-	44.8	47.4	48.0	59.0	63.4	65.2	50.7	50.6	43.6
VG	58.1	58.3	39.7	52.6	54.6	43.1	58.5	-	-	65.5	68.8	65.2	50.1	51.3	43.6
COCOQA	30.1	35.5	39.7	35.1	40.4	43.1	29.1	33.1	48.0	75.8	-	-	33.3	37.5	43.6
VQA2⁻	48.8	50.8	39.7	55.2	55.3	43.1	47.3	49.1	48.0	60.3	64.9	65.2	53.8	-	-



Analysis:

- VG (COCOQA) generalizes the best (worst).
- COCOQA is improved the most.

Conclusion:

- Our method is robust and widely applicable.

[6] W.-L. Chao, H. Hu, and F. Sha. Being negative but constructively. In NAACL, 2018
[36] B. Sun, J. Feng, and K. Saenko. Return of frustratingly easy domain adaptation. In AAAI, 2016
[41] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell. Adversarial discriminative domain adaptation. In CVPR, 2017