Impact of ImageNet Model Selection on Domain Adaptation

Youshan Zhang & Brian D. Davison

03/01/2020

LEHIGH UNIVERSITY
Computer Science & Engineering

Outline

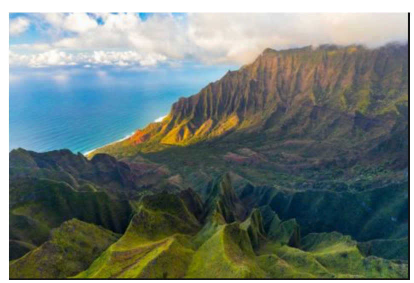
1. Introduction
2. Problem
3. State-of-the-art methods
4. Methods
5. Datasets & Results
6. Conclusion & Future work

1 Introduction

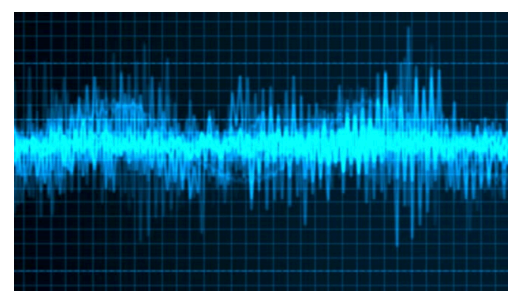
Background

- Big data & different types of data
- Original elementary form & not labeled
- Manually label data: time-consuming & expensive
- Model reuse based on labeled data
  - Conventional machine learning: lower accuracy

Text



Image

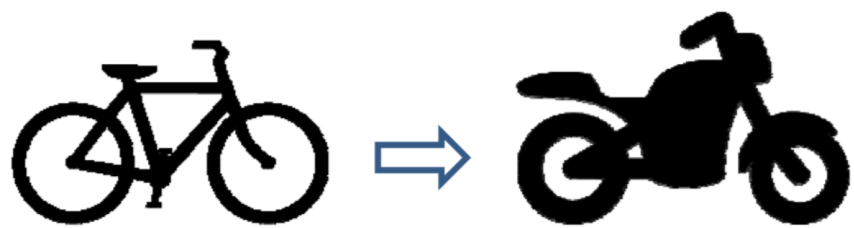


Audio

1 What is domain adaptation (DA)?

- Human: previous experience and knowledge for reasoning & learning

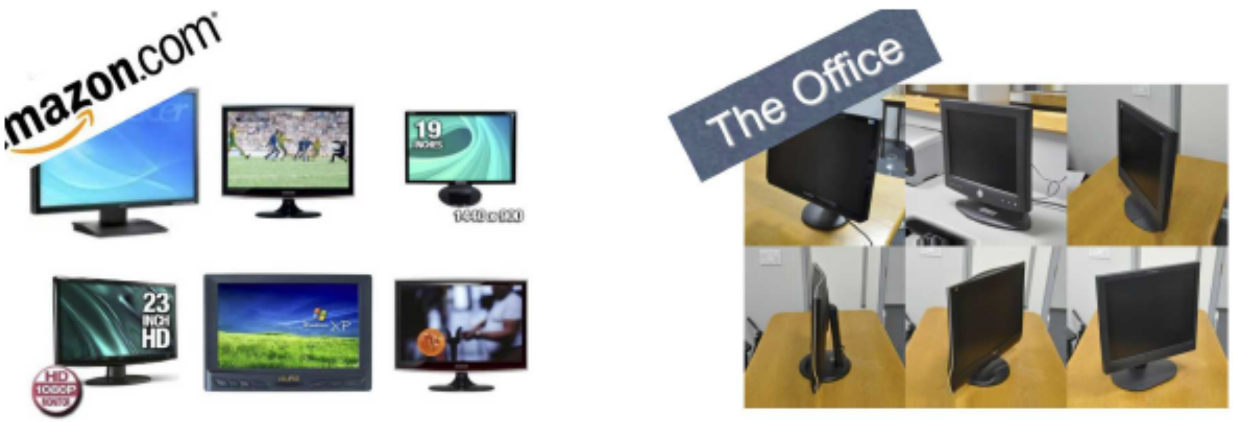
- Machine: apply knowledge from other fields into current applications



- Key idea:
  - ➢ How to find similar domain knowledge for transferring?

Why DA ?

- Label data: time-consuming & expensive
- Train from scratch: tedious
- Design customized model: complex
- Data shift/bias



TRAIN → **TEST**

Source Domain **Target Domain**

Outline

1. Introduction
2. **Problem**
3. State-of-the-art methods
4. Methods
5. Datasets & Results
6. Conclusion & Future work

2

Problem

- Source domain ($\mathcal{D}_S$) (labeled)

$$X_S = \{ (x_i, y_i) \mid i = 1, 2, \ldots , n_s\}$$

- Target domain ($\mathcal{D}_T$) (unlabeled)

$$X_T = \{ (x_j, ?) \mid j = 1, 2, \cdots , n_t\}$$

- Objective
  - Train classification model on source domain and improve the accuracy on the target

Outline

1. Introduction
2. Problem
3. **State-of-the-art methods**
4. Methods
5. Datasets & Results
6. Conclusion & Future work

3

State-of-the-art methods

- Traditional methods
  - Feature selection[Blitzer et al., 2006; Long et al., 2014]
  - Subspace learning[Gopalan et al., 2011; Gonget al., 2012; Zhang et al., 2019b]
  - Distribution adaptation[Panet al., 2011; Jiang et al., 2017; Wang et al., 2018]
- Deep learning based methods
  - Discrepancy based[Tzeng et al., 2014; Long et al.,2015; Ghifary et al., 2015]
  - Reconstruction based[Bousmalis et al., 2016]
  - Adversarial learning based[Ganinet al., 2016; Tzeng et al., 2017; Liu et al., 2019]

3

Limitations

- More or less rely on the backbone networks
- Not explore other ImageNet models
- Not know which is the best layer

Outline

1 Introduction

2 Problem

3 State-of-the-art methods

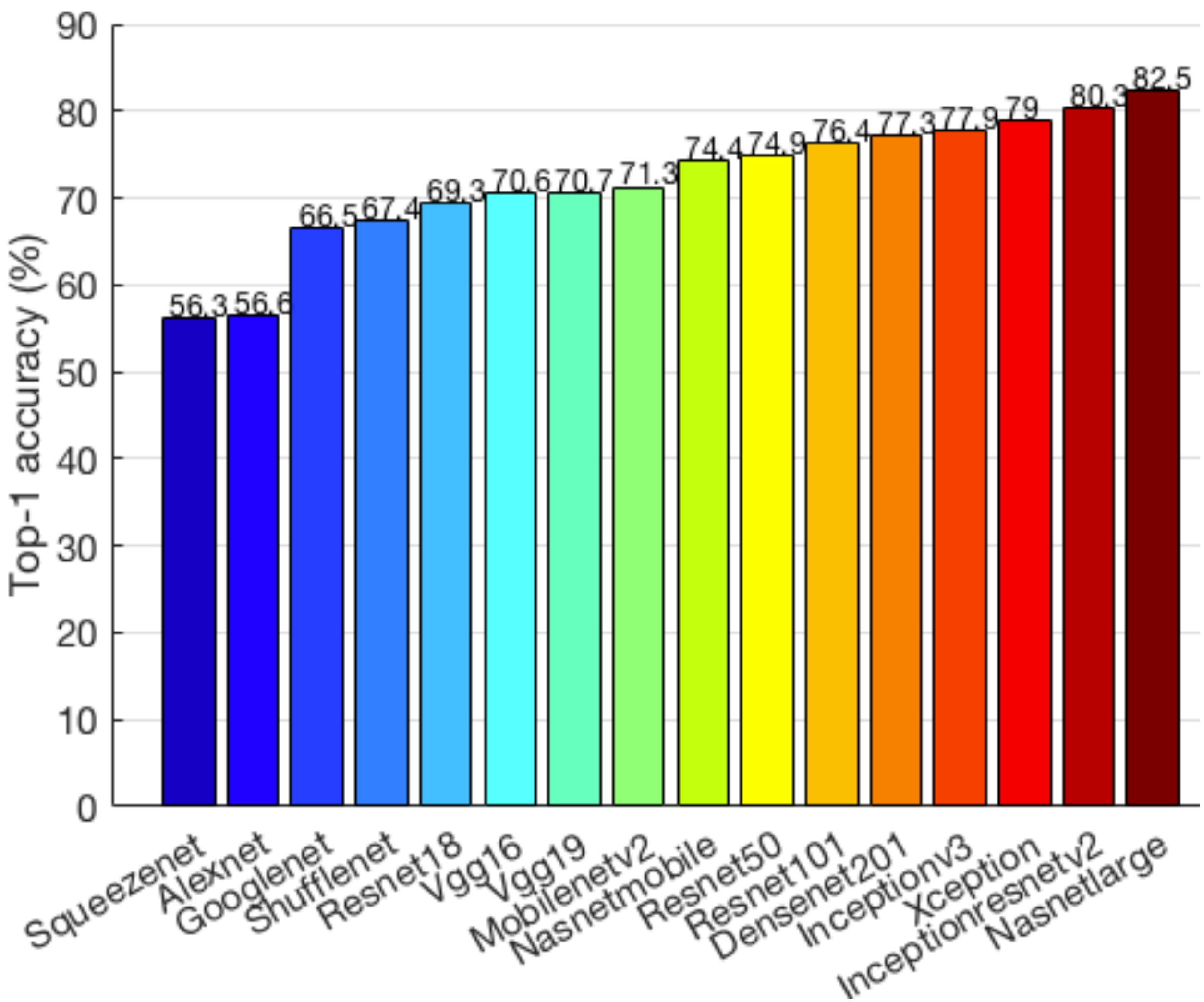
4 **Methods**

5 Datasets & Results

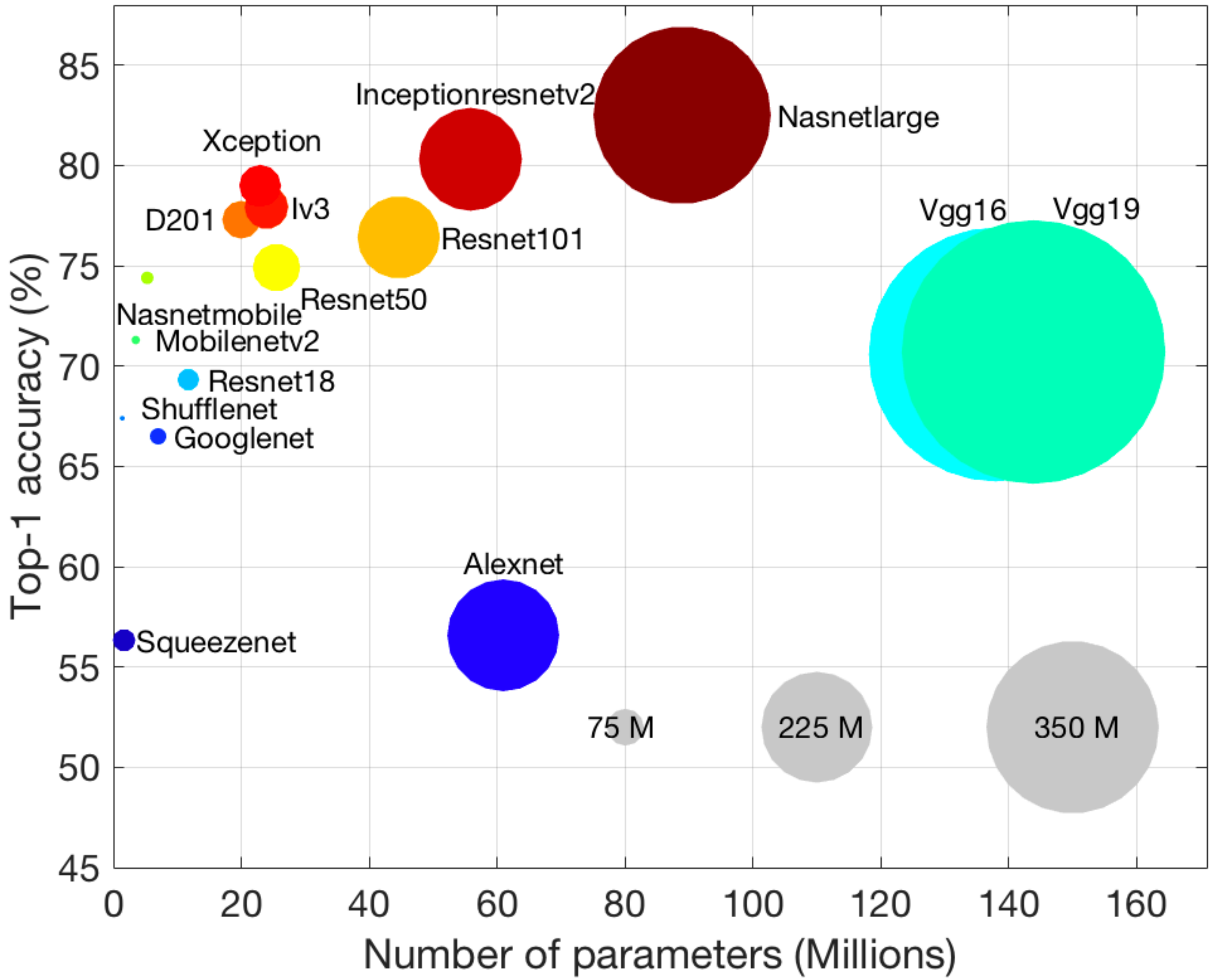
6 Conclusion & Future work

4

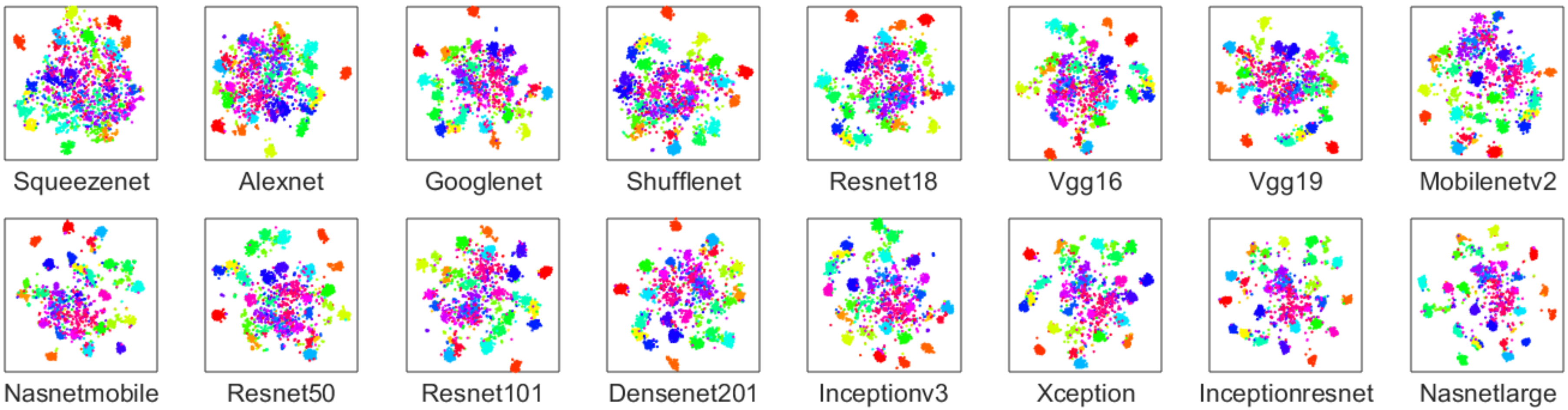
ImageNet Models

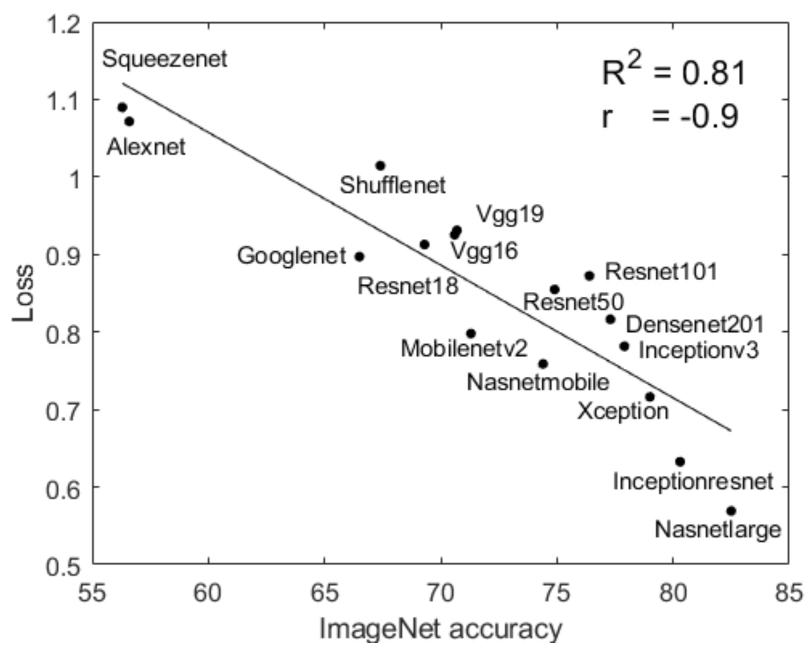


ImageNet Models



4 Extracted features visualization





4 Domain adaptation methods

- Support vector machines (SVM) & 1-nearest neighbor (1NN)
- Geodesic Flow Kernel (GFK) & Geodesic sampling on manifolds (GSM)
- CORrelation Alignment (CORAL)
- Transfer Joint Matching (TJM)
- Balanced distribution adaptation(BDA) & Joint distribution alignment (JDA) & Joint Geometrical and Statistical Alignment (JGSA) & Adaptation Regularization (ARTL) & Manifold Embedded Distribution Alignment (MEDA) & Modified Distribution Alignment (MDA)

4 Significance analysis

- Correlation coefficient

$$r(A,B) = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n A_{mn} - \bar{A})^2 (\sum_m \sum_n B_{mn} - \bar{B})^2}}$$

- Coefficient of determination

$$R^2 = 1 - \frac{\text{Unexplained variation}}{\text{Total variation}} = 1 - \frac{\sum_{i=1}^{N} S_{residual}}{\sum_{i=1}^{N} S_{total}}$$

$$S_{residual} = (y_i - y_i')^2, S_{total} = (y_i - \bar{y})^2$$

- The higher the better

Outline

1. Introduction
2. Problem
3. State-of-the-art methods
4. Methods
5. **Datasets & Results**
6. Conclusion & Future work

5

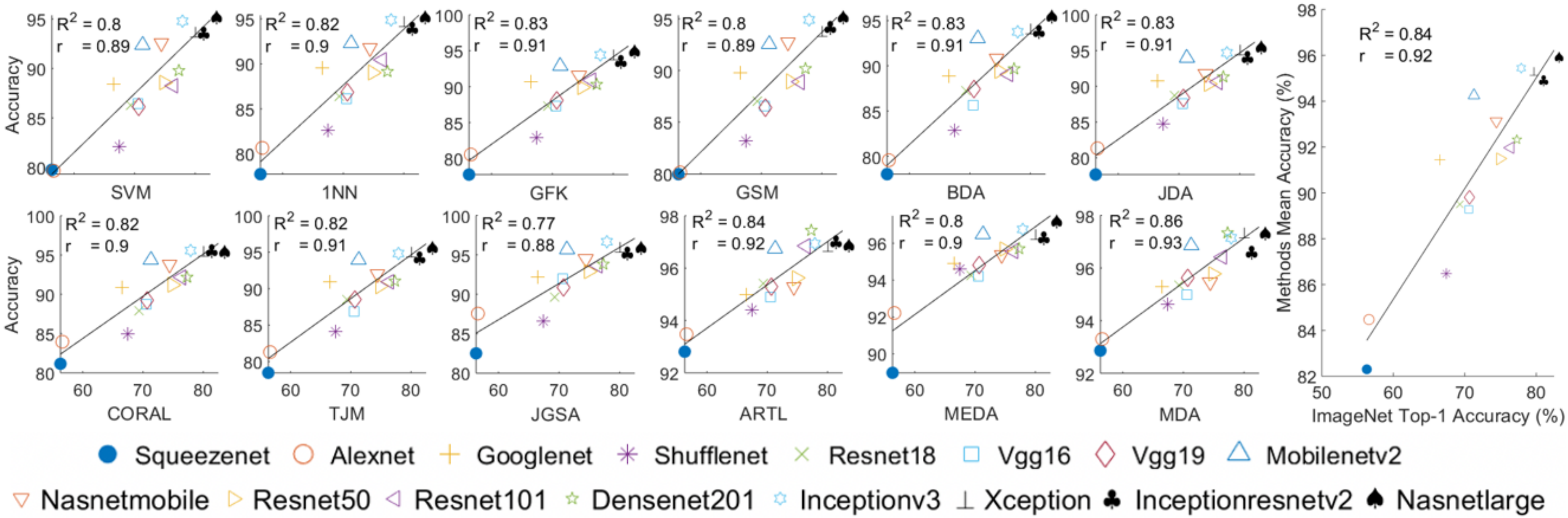
Datasets & Results

Datasets

Table 1: Statistics of extracted IR features of three datasets

Dataset	# Sample	# Feature	# Class	Domains
Office + Caltech-10	2533	1000	10	A, C, W, D
Office-31	4110	1000	31	A, W, D
Office-Home	15588	1000	65	A, C, P, R

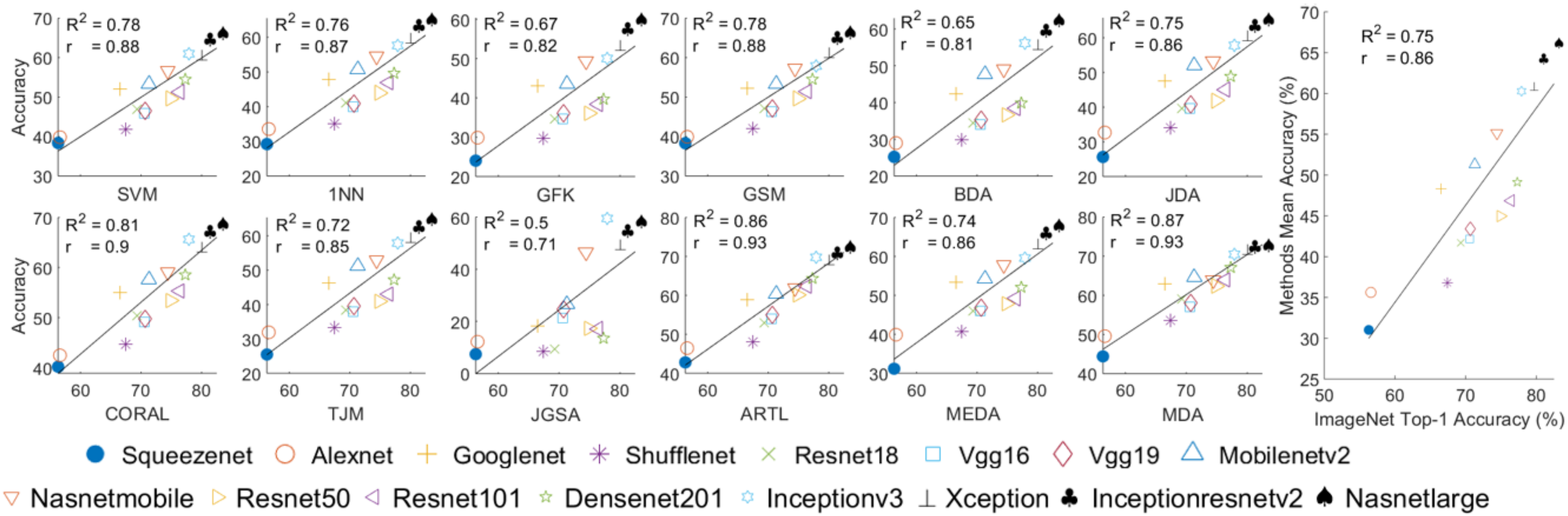
Results



Office + Caltech 10

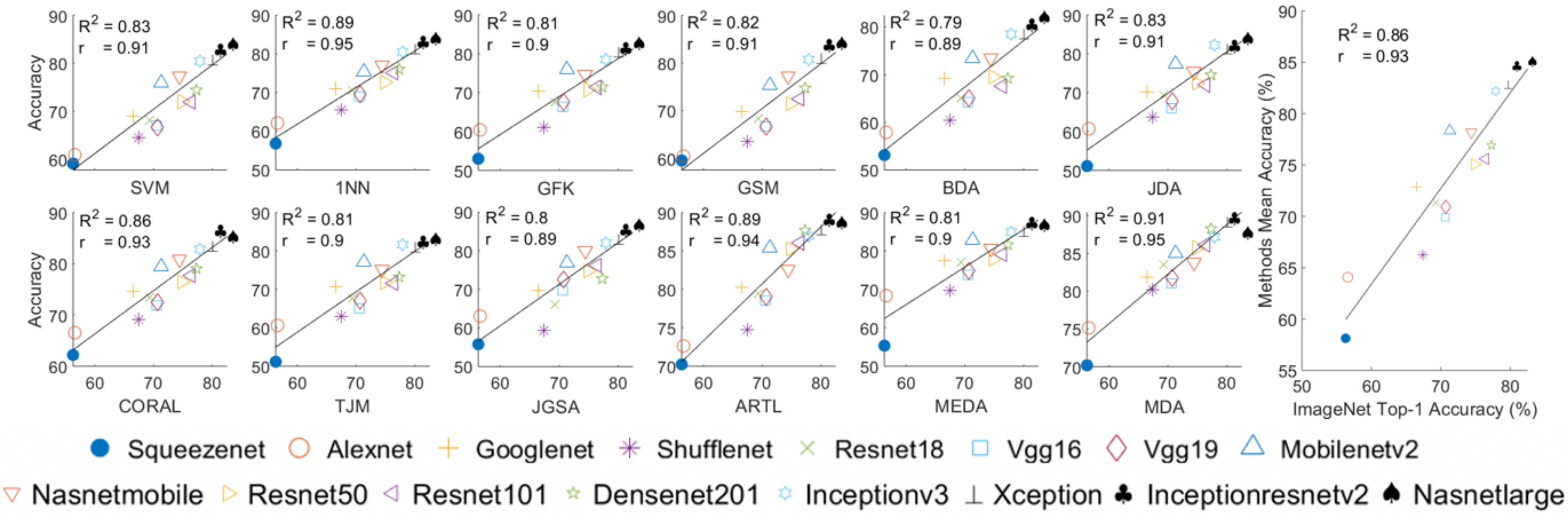
5

Results



Office-Home

5 Results



Office31

5 Classification Accuracy

Table 2: Accuracy (%) on Office + Caltech-10 datasets

Task	C → A	C → W	C → D	A → C	A → W	A → D	W → C	W → A	W → D	D → C	D → A	D → W	Average
SVM	94.7	97.3	99.4	93.3	90.5	92.4	93.9	95.4	**100**	94.2	94.4	99.0	95.4
1NN	95.7	96.3	95.5	93.6	91.5	95.5	93.7	95.7	**100**	93.5	94.8	98.3	95.3
GFK [11]	94.8	96.6	94.9	92.4	92.5	94.9	93.6	95.2	**100**	94.2	94.4	98.3	95.2
GSM [54]	95.6	96.3	98.1	93.9	90.2	93.0	93.9	95.5	**100**	94.4	94.4	99.0	95.4
BDA [46]	95.7	95.6	96.8	92.8	96.6	94.9	93.5	95.8	**100**	93.3	95.8	96.3	95.6
JDA [26]	95.3	96.3	96.8	93.9	95.9	95.5	93.5	95.7	**100**	93.3	95.5	96.9	95.7
CORAL [37]	95.6	96.3	98.1	95.2	89.8	94.3	93.9	95.7	**100**	94.0	96.2	98.6	95.6
TJM [27]	95.7	96.6	95.5	93.2	95.9	97.5	93.4	95.7	**100**	93.5	95.6	96.9	95.8
JGSA [49]	95.2	97.6	96.8	95.2	93.2	95.5	94.6	95.2	**100**	94.9	96.1	99.3	96.1
ARTL [25]	95.7	97.6	97.5	94.6	98.6	**100**	94.6	96.1	**100**	93.5	95.8	99.3	96.9
MEDA [48]	**96.0**	**99.3**	98.1	94.2	99.0	100	94.6	96.5	**100**	94.1	96.1	99.3	97.3
MDA [53]	**96.0**	**99.3**	**99.4**	**94.2**	99.0	**100**	**94.6**	**96.5**	**100**	**94.2**	**96.1**	99.3	**97.4**
DAN [23]	92.0	90.6	89.3	84.1	91.8	91.7	81.2	92.1	**100**	80.3	90.0	98.5	90.1
DDC [43]	91.9	85.4	88.8	85.0	86.1	89.0	78.0	83.8	**100**	79.0	87.1	97.7	86.1
DCORAL [38]	89.8	97.3	91.0	91.9	**100**	90.5	83.7	81.5	90.1	88.6	80.1	92.3	89.7
RTN [28]	93.7	96.9	94.2	88.1	95.2	95.5	86.6	92.5	**100**	84.6	93.8	99.2	93.4
MDDA [33]	93.6	95.2	93.4	89.1	95.7	96.6	86.5	94.8	**100**	84.7	94.7	**99.4**	93.6

5 Classification Accuracy

Table 3: Accuracy (%) on Office-Home datasets

Task	Ar → Cl	Ar → Pr	Ar → Rw	Cl → Ar	Cl → Pr	Cl → Rw	Pr → Ar	Pr → Cl	Pr → Rw	Rw → Ar	Rw → Cl	Rw → Pr	**Average**
SVM	47.8	76.1	79.2	61.7	70.2	69.5	64.4	48.7	79.5	70.6	49.1	82.1	66.6
1NN	46.4	71.7	77	63.9	69.6	70.4	65.5	46.8	76.0	71.4	48.5	78.7	65.5
GFK [11]	39.6	66.0	72.5	55.7	66.4	64.0	58.4	42.5	73.3	66.0	44.1	76.1	60.4
GSM [54]	47.6	76.4	79.5	62.2	69.7	69.2	65.1	49.5	79.8	71.0	49.6	82.1	66.8
BDA [46]	43.3	69.8	74.1	58.7	66.3	67.7	60.6	46.3	75.3	67.3	48.7	77.0	62.9
JDA [26]	47.4	72.8	76.1	60.7	68.6	70.5	66.0	49.1	76.4	69.6	52.5	79.7	65.8
CORAL [37]	48.0	78.7	80.9	65.7	74.7	75.5	68.4	49.8	80.7	73.0	50.1	82.4	69.0
TJM [27]	47.6	72.3	76.1	60.7	68.6	71.1	64.0	49.0	75.9	68.6	51.2	79.2	65.4
JGSA [49]	42.9	69.5	71.2	50.1	63.0	63.3	55.6	42.6	71.8	60.8	42.1	74.6	59.0
ARTL [25]	53.5	80.2	81.6	71.5	79.9	78.3	**73.1**	56.1	**82.9**	**75.9**	**57.1**	83.7	72.8
MEDA [48]	48.5	74.5	78.8	64.8	76.1	75.2	67.4	49.1	79.7	72.2	51.7	81.5	68.3
MDA [53]	**54.8**	**81.2**	**82.3**	**71.9**	**82.9**	**81.4**	71.1	**53.8**	82.8	75.5	55.3	**86.2**	**73.3**
DCORAL [38]	32.2	40.5	54.5	31.5	45.8	47.3	30.0	32.3	55.3	44.7	42.8	59.4	42.8
RTN [28]	31.3	40.2	54.6	32.5	46.6	48.3	28.2	32.9	56.4	45.5	44.8	61.3	43.5
DAH [44]	31.6	40.8	51.7	34.7	51.9	52.8	29.9	39.6	60.7	45.0	45.1	62.5	45.5
MDDA [33]	35.2	44.4	57.2	36.8	52.5	53.7	34.8	37.2	62.2	50.0	46.3	66.1	48.0
DAN [23]	43.6	57.0	67.9	45.8	56.5	60.4	44.0	43.6	67.7	63.1	51.5	74.3	56.3
DANN [10]	45.6	59.3	70.1	47.0	58.5	60.9	46.1	43.7	68.5	63.2	51.8	76.8	57.6
JAN [29]	45.9	61.2	68.9	50.4	59.7	61.0	45.8	43.4	70.3	63.9	52.4	76.8	58.3
CDAN-RM [24]	49.2	64.8	72.9	53.8	62.4	62.9	49.8	48.8	71.5	65.8	56.4	79.2	61.5
CDAN-M [24]	50.6	65.9	73.4	55.7	62.7	64.2	51.8	49.1	74.5	68.2	56.9	80.7	62.8

5

Classification Accuracy

Table 4: Accuracy (%) on Office 31 datasets

Task	A → W	A → D	W → A	W → D	D → A	D → W	**Average**
SVM	81.5	80.9	73.4	96.6	70.6	95.1	83.0
1NN	80.3	81.1	71.8	99.0	71.3	96.4	83.3
GFK [11]	78.1	78.5	71.7	98.0	68.9	95.2	81.7
GSM [54]	84.8	82.7	73.5	96.6	70.9	95.0	83.9
BDA [46]	77.0	79.3	70.3	97.0	68.0	93.2	80.8
JDA [26]	79.1	79.7	72.9	97.4	71.0	94.2	82.4
CORAL [37]	88.9	87.6	74.7	99.2	73.0	96.7	86.7
TJM [27]	79.1	81.1	72.9	96.6	71.2	94.6	82.6
JGSA [49]	81.1	84.3	76.5	99.0	75.8	97.2	85.7
ARTL [25]	92.5	91.8	76.9	99.6	77.1	97.5	89.2
MEDA [48]	90.8	91.4	74.6	97.2	75.4	96.0	87.6
MDA [53]	**94.0**	**92.6**	77.6	99.2	78.7	96.9	**89.8**
DAN [23]	80.5	78.6	62.8	99.6	63.6	97.1	80.4
RTN [28]	84.5	77.5	64.8	99.4	66.2	96.8	81.6
DANN [10]	82.0	79.7	67.4	99.1	68.2	96.8	81.6
ADDA [42]	86.2	77.8	68.9	98.4	69.5	96.2	82.9
CAN [50]	81.5	65.9	**98.2**	85.5	**99.7**	63.4	82.4
JDDA [3]	82.6	79.8	66.7	99.7	57.4	95.2	80.2
JAN [29]	85.4	84.7	70.0	**99.8**	68.6	**97.4**	84.3
GCAN [30]	82.7	76.4	62.6	**99.8**	64.9	97.1	80.6

5

Best feature extraction layer

Task	Output	Softmax	LFC	P_LFC
Squeezenet [17]	42.0	42.0	**44.4**	-
Alexnet [21]	43.0	43.0	49.6	**50.4**
Googlenet [40]	53.0	53.0	62.9	**64.2**
Shufflenet [51]	45.9	45.9	53.5	**54.7**
Resnet18 [15]	49.5	49.5	59.2	**62.0**
Vgg16 [36]	47.8	47.8	57.1	**58.3**
Vgg19 [36]	48.4	48.4	58.0	**59.4**
Mobilenetv2 [35]	52.4	52.4	52.4	**64.7**
Nasnetmobile [55]	52.8	52.8	63.8	**64.6**
Resnet50 [15]	50.0	50.0	62.4	**62.5**
Resnet101 [15]	51.2	51.2	63.9	**64.7**
Densenet201 [16]	54.3	54.3	67.1	**69.5**
Inceptionv3 [41]	57.4	57.4	69.7	**70.4**
Xception [4]	59.4	59.4	72.0	**72.3**
Inceptionresnetv2 [39]	60.1	60.1	72.8	**73.8**
Nasnetlarge [55]	60.6	60.6	73.3	**73.6**

5

Take home messages

- Features from a higher-performing ImageNet-trained model are more valuable than those from a lower-performing model for unsupervised domain adaptation

- The layer prior to the last fully connected layer is the best layer for feature extraction

6

Conclusion & future work

- We are the first to examine how features from many different ImageNet models affect domain adaptation

- Search the best architecture for feature extraction
- Feature fusion

6

Thank you!

Questions?