### 4.1 Results

Following the fully transductive evaluation, we summarize the results in Table 1 and Table 2. Table 1 summarizes the results on the object recognition task using office dataset whereas Table 2 summarizes the digit classification task on MNIST and SVHN.

Table 1: Accuracy of our method and the state-of-the-art algorithms on Office dataset.

| SOURCE<br>TARGET | AMAZON<br>WEBCAM | D-SLR<br>WEBCAM | WEBCAM<br>D-SLR | WEBCAM<br>AMAZON | AMAZON<br>D-SLR | D-SLR<br>AMAZON |
|---|---|---|---|---|---|---|
| GFK [12] | .398 | .791 | .746 | .371 | .379 | .379 |
| SA* [9] | .450 | .648 | .699 | .393 | .388 | .420 |
| DLID [6] | .519 | .782 | .899 | - | - | - |
| DDC [33] | .618 | .950 | .985 | .522 | .644 | .521 |
| DAN [20] | .685 | .960 | .990 | .531 | .670 | .540 |
| BACKPROP [11] | .730 | **.964** | **.992** | .536 | .728 | .544 |
| SOURCE ONLY | .642 | .961 | .978 | .452 | .668 | .476 |
| OUR METHOD (K-NN ONLY) | .727 | .952 | .915 | .575 | .791 | .521 |
| OUR METHOD (NO REJECT) | .804 | .962 | .989 | .625 | .839 | .567 |
| OUR METHOD (FULL) | **.811** | **.964** | **.992** | **.638** | **.841** | **.583** |

Tables 1&2 show results on object recognition and digit classification tasks covering all adaptation scenarios. Our experiments show that our proposed method outperforms all state-of-the-art algorithms. Moreover, the increase in the accuracy is rather significant when there is a large domain difference such as MNIST↔MNIST-M, MNIST↔SVHN, Amazon↔Webcam and Amazon↔D-SLR. Our hypothesis is that the state-of-the-art algorithms such as [11] are seeking features invariant to the domains whereas we seek an explicit similarity metric explaining both differences and similarities of domains. In other words, instead of seeking an invariance, we seek an equivariance.

Table 2: Accuracy on the digit classification task.

| SOURCE<br>TARGET | M-M<br>MNIST | MNIST<br>M-M | SVHN<br>MNIST | MNIST<br>SVHN |
|---|---|---|---|---|
| SA* [9] | .523 | .569 | .593 | .211 |
| BP [11] | .732 | .766 | .738 | .289 |
| SOURCE ONLY | .483 | .522 | .549 | .162 |
| OUR METHOD(K-NN ONLY) | .805 | .795 | .713 | .158 |
| OUR METHOD(NO REJECT) | .835 | .855 | .774 | .323 |
| OUR METHOD(FULL) | **.839** | **.867** | **.788** | **.403** |

Table 2 further suggests that our algorithm is the only one which can successfully perform adaptation from MNIST to SVHN. Clearly the features which are learned from MNIST cannot generalize to SVHN since the SVHN has concepts like color and occlusion which are not available in MNIST. Hence, our algorithm learns SVHN specific features by enforcing accurate transduction in the adaptation.

Another interesting conclusion is the asymmetric results. For example, adapting webcam to Amazon and adapting Amazon to webcam yield very different accuracies. The similar asymmetry exists in MNIST and SVHN as well. This observation validates the importance of an asymmetric modeling.

To evaluate the importance of joint labelling and reject option, we compare our method with self baselines. Our self-baselines are versions of our algorithm not using the reject option (**no reject**) and the version using neither reject option nor joint labelling (**k-NN only**). Results on both experiments suggest that joint labelling and the reject option are both crucial for successful transduction. Moreover, the reject option is more important when the domain shift is large (e.g. MNIST→SVHN). This is expected since transduction under a large shift is more likely to fail a situation that can be prevented with reject option.

#### 4.1.1 Qualitative Analysis

To further study the learned representations and the similarity metric, we performed a series of qualitative analysis in the form of nearest neighbor and tSNE[34] plots.

Figure 1 visualizes example target images from MNIST and their corresponding source images. First of all, our experimental analysis suggests that MNIST and SVHN are the two domains with the largest difference. Hence, we believe MNIST↔SVHN is a very challenging set-up and despite the huge