Supplementary Material for
Multi-view Classification Using Hybrid Fusion and Mutual Distillation

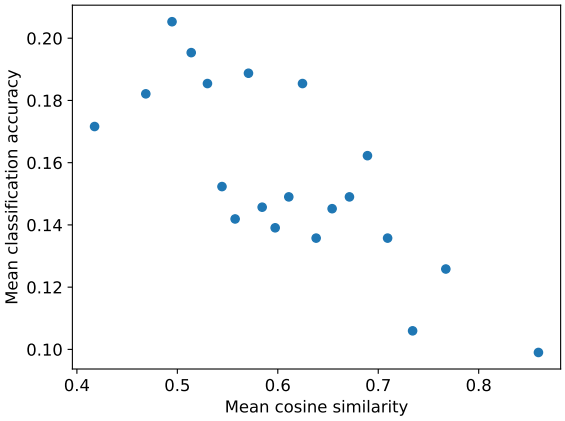

Figure 1. Accuracy vs paired image similarity. This plot shows the mean classification accuracy as a function of the image pair similarity. There is a strong negative correlation between image similarity and classification accuracy ($r = -.775$)

Figure 2. Examples of an incorrectly classified high-similarity pair (top row) and correctly classified low-similarity pair (bottom).

We include additional experimental results that were omitted due to space constraints.

A. Paired-Image Similarity Analysis

In the classic cross-view case typified by medical image analysis, the assumption is that the images in the pair contribute orthogonal information, if not explicitly, at least figuratively. In the more general case, especially with hotel images, the amount of overlap between a given pair of images can vary substantially. It is reasonable to assume that "new" information is the key to improved multi-view performance and that the higher the dissimilarity between a pair of images, the more likely that multi-view processing will provide a benefit.

Here, we consider the subset of test examples where each image is incorrectly classified under the single-view regime. To measure image (dis)similarity, we compute the cosine similarity between the InceptionV3 features for each image in the pair. We subdivide the range of paired image similarities into 20 evenly-sized bins and compute the mean accuracy for the image pairs assigned to the bin when classified under the multi-view regime. The result of this analysis is shown in Figure 1.

As expected, there is a strong negative correlation between image similarity and classification accuracy ($r = -.775$), indicating that image pairs where the constituent images are visually similar do not benefit as much from multi-view processing as those that introduce complementary visual information. In the context of hotel room images, we observe that paired examples containing similar objects (e.g., bed, carpet, and chair) fall into this category, as shown in Figure 2

B. CheXpert Evaluation

Experimental Procedure We follow the procedure outlined in [49] when training and evaluating on the CheXpert cross-view medical classification dataset. The model is optimized for each of the 13 detection tasks simultaneously and the per-task loss is weighted equally in the overall ob-

jective function. When training, tasks that are assigned the "missing" label are disregarded, and the loss is computed using a predicted 3-class distribution over the "negative", "positive", and "uncertain" labels. At inference, the tasks that are assigned the "uncertain" label are removed, and the 2-class softmax distribution is predicted using the logits from the "negative" and "postive" labels. The model checkpoint that achieves the highest AUC-ROC for a given task on the validation set is saved and then used to evaluate that task on the test set.

Results Table 1 includes the expanded per-task AUC-ROC. MV-HFMD outperforms the baselines on all tasks except for edema and fracture detection. Additionally, frontal and lateral view results are presented in Table 2 and Table 3, respectively, for the methods that output predictions for each individual view in addition to the entire collection. Compared to the baselines, MV-HFMD achieves better multi-view accuracy for a given task relative to the max single-view score for that task. Moreover, MVC-Net has five tasks where higher accuracy is achieved on either of the individual views compared to the multi-view (atelectasis, enlarged cardiomed., fracture, lung lesion, and pneumonia). For MV-HFMD with unshared weights, there are only two such tasks (lung lesion and lung opacity).

Task	MVCNN	CVT*	MVC-NET*	GVCNN	TMC*	MVT	MV-HFMD	MV-HFMD*
Overall	.815 ± .004	.813 ± .003	.813 ± .005	.805 ± .003	.802 ± .002	.816 ± .003	.835 ± .003	**.845 ± .002**
Atelectasis	.781 ± .018	.780 ± .033	.809 ± .024	.802 ± .013	.790 ± .029	.805 ± .024	.828 ± .025	**.846 ± .017**
Cardiomegaly	.918 ± .003	.919 ± .005	.917 ± .007	.907 ± .004	.913 ± .003	.927 ± .004	.927 ± .002	**.932 ± .002**
Consolidation	.853 ± .006	.860 ± .006	.852 ± .014	.844 ± .012	.843 ± .013	.880 ± .005	.881 ± .007	**.883 ± .002**
Edema	.880 ± .003	.881 ± .002	.874 ± .003	.873 ± .009	**.882 ± .004**	.876 ± .005	.875 ± .005	.880 ± .002
Enlarged Cardiomed.	.802 ± .005	.804 ± .007	.800 ± .011	.796 ± .002	.799 ± .007	.805 ± .006	.814 ± .005	**.823 ± .002**
Fracture	**.769 ± .025**	.769 ± .015	.744 ± .009	.730 ± .005	.765 ± .003	.739 ± .020	.757 ± .010	.767 ± .013
Lung Lesion	.720 ± .011	.729 ± .013	.717 ± .013	.717 ± .006	.729 ± .008	.746 ± .007	.746 ± .007	**.754 ± .007**
Lung Opacity	.799 ± .003	.795 ± .009	.806 ± .008	.799 ± .006	.808 ± .004	.798 ± .003	.817 ± .003	**.818 ± .001**
Pleural Effusion	.954 ± .001	.955 ± .001	.955 ± .003	.948 ± .003	.956 ± .001	.956 ± .001	**.964 ± .001**	.962 ± .000
Pleural Other	.800 ± .036	.759 ± .034	.770 ± .011	.766 ± .049	.631 ± .047	.753 ± .019	.783 ± .035	**.821 ± .015**
Pneumonia	.737 ± .013	.732 ± .025	.731 ± .005	.734 ± .008	.710 ± .027	.734 ± .008	.759 ± .011	**.774 ± .014**
Pneumothorax	.851 ± .009	.850 ± .009	.846 ± .003	.844 ± .002	.857 ± .009	.856 ± .009	**.874 ± .002**	.873 ± .004
Support Devices	.734 ± .021	.733 ± .017	.752 ± .021	.704 ± .011	.742 ± .027	.735 ± .029	.834 ± .016	**.857 ± .005**

Table 1. Performance on cross-view chest x-ray classification. The table shows the mean and standard deviation AUC-ROC across 13 classification tasks repeated over four training runs. * indicates unshared weights for input views.

Task	MVC-NET*	TMC*	MV-HFMD	MV-HFMD*
Overall	.800 ± .003	.786 ± .001	.822 ± .004	**.828 ± .002**
Atelectasis	.787 ± .027	.749 ± .028	.831 ± .030	**.837 ± .012**
Cardiomegaly	.914 ± .009	.907 ± .010	.929 ± .003	**.931 ± .001**
Consolidation	.837 ± .012	.822 ± .014	.871 ± .010	**.876 ± .005**
Edema	.858 ± .002	.865 ± .009	.863 ± .004	**.866 ± .006**
Enlarged Cardiomed.	.801 ± .012	.786 ± .013	.816 ± .004	**.817 ± .002**
Fracture	**.748 ± .018**	.730 ± .022	.738 ± .011	.746 ± .022
Lung Lesion	.728 ± .008	.719 ± .006	**.764 ± .008**	.757 ± .011
Lung Opacity	.796 ± .009	.806 ± .005	.818 ± .004	**.819 ± .005**
Pleural Effusion	.933 ± .003	.934 ± .004	**.947 ± .001**	**.947 ± .001**
Pleural Other	**.759 ± .033**	.683 ± .031	.751 ± .039	.755 ± .024
Pneumonia	.700 ± .007	.677 ± .029	.727 ± .017	**.746 ± .014**
Pneumothorax	.826 ± .004	.834 ± .013	.852 ± .002	**.855 ± .008**
Support Devices	.708 ± .006	.703 ± .020	.774 ± .010	**.816 ± .020**

Table 2. Performance on chest x-ray classification using the frontal view only.

Task	MVC-NET*	TMC*	MV-HFMD	MV-HFMD*
Overall	.796 ± .007	.774 ± .004	.815 ± .004	**.818 ± .004**
Atelectasis	.822 ± .014	.805 ± .020	.810 ± .010	**.830 ± .006**
Cardiomegaly	.890 ± .009	.876 ± .007	.898 ± .004	**.906 ± .004**
Consolidation	.839 ± .016	.832 ± .022	**.862 ± .008**	**.862 ± .003**
Edema	**.861 ± .003**	.856 ± .005	.855 ± .004	**.861 ± .003**
Enlarged Cardiomed.	.779 ± .008	.776 ± .005	.786 ± .009	**.801 ± .003**
Fracture	.746 ± .014	**.754 ± .009**	.710 ± .021	.743 ± .030
Lung Lesion	.703 ± .009	.713 ± .016	**.719 ± .011**	.700 ± .030
Lung Opacity	.781 ± .002	.780 ± .004	.787 ± .003	**.792 ± .004**
Pleural Effusion	.944 ± .004	.946 ± .005	**.952 ± .001**	**.952 ± .001**
Pleural Other	.700 ± .052	.559 ± .052	**.805 ± .040**	.757 ± .032
Pneumonia	**.740 ± .007**	.652 ± .051	.728 ± .012	.739 ± .015
Pneumothorax	.829 ± .005	.820 ± .012	.856 ± .003	**.860 ± .005**
Support Devices	.720 ± .043	.696 ± .042	.824 ± .013	**.837 ± .008**

Table 3. Performance on chest x-ray classification using the lateral view only.