An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification

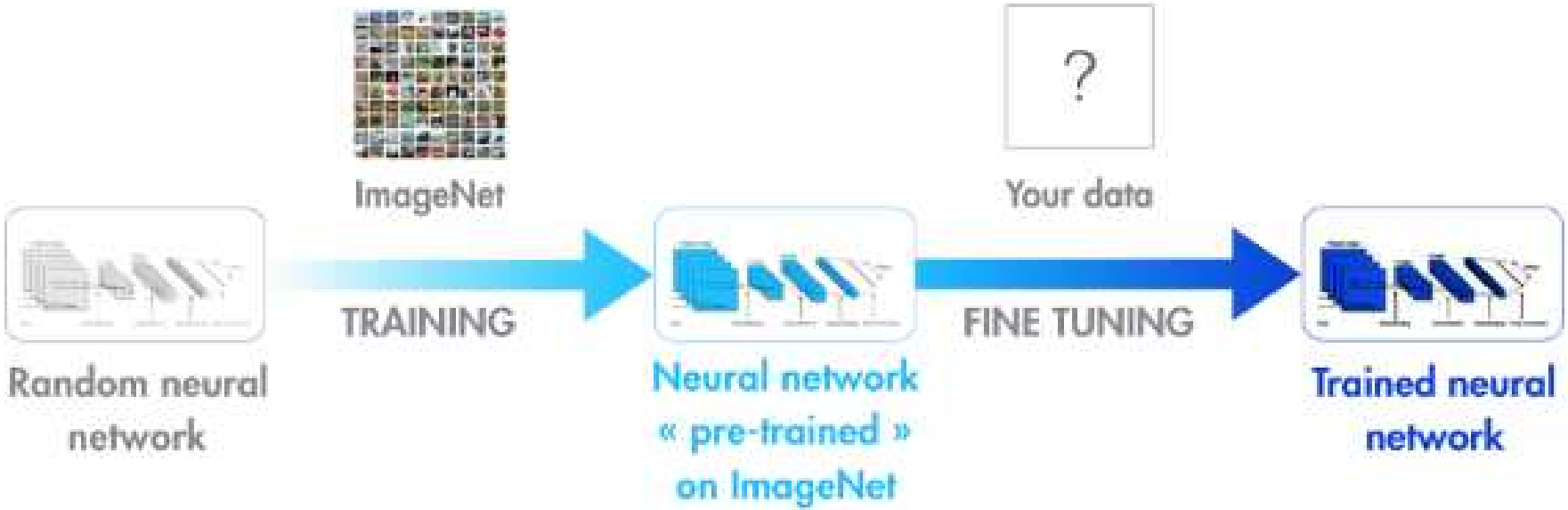
Ashnil Kumar, et al. IEEE J. Biomed. and Health Informatics, 1 Jan. 2017

2018. 7

Contents

- Neural Learning transfer model (Medical Image Classification)
- Neural Learning transfer(Ensemble)
- Data Augmentation
- Neural networks(AlexNet/GoogleNet)
- Result
- Discussion & Conclusion

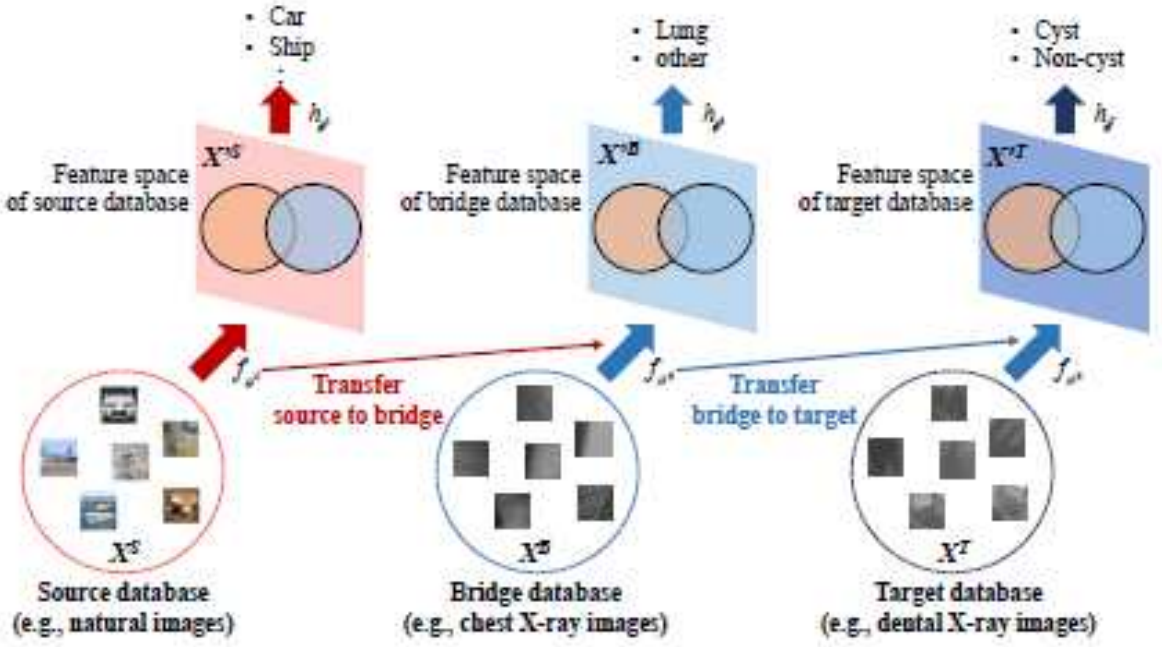
Neural Transfer learning model



https://arxiv.org/abs/1602.03409

Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning

Transfer learning with Modality-bridge



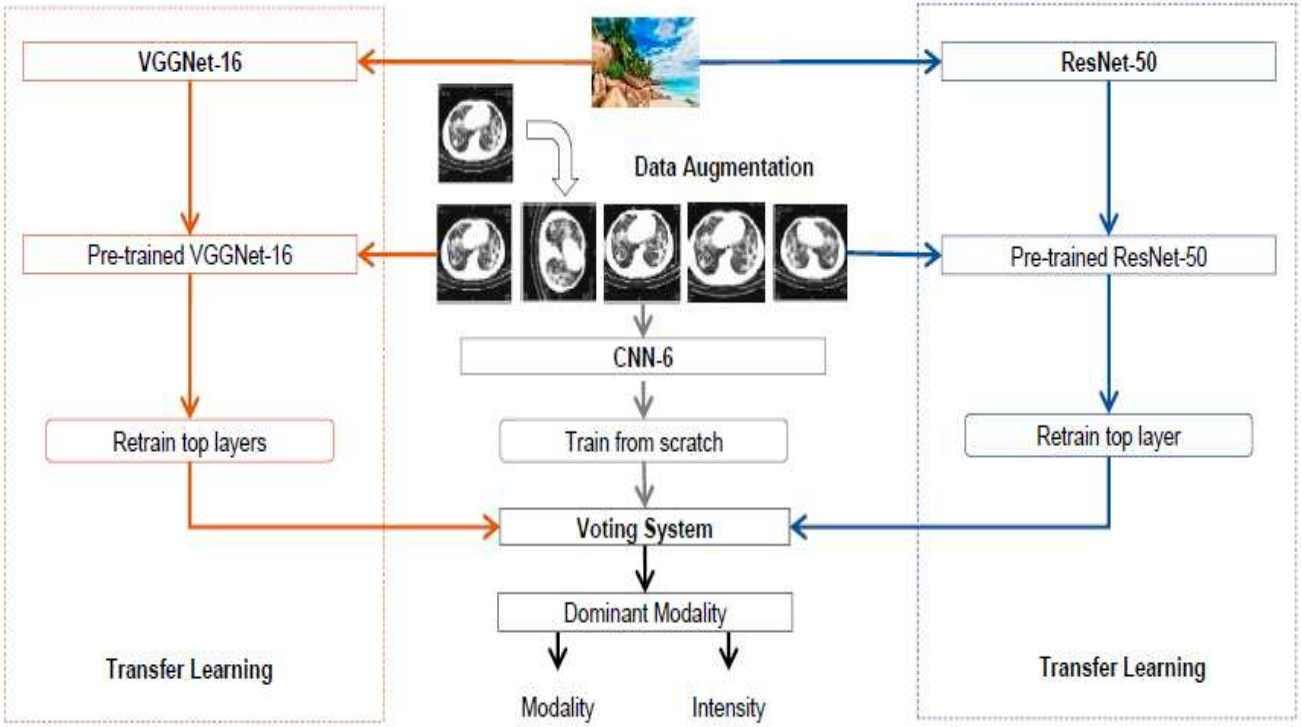
Hak Gu Kim., et al., Modality-bridge Transfer Learning for Medical Image Classification, https://arxiv.org ›2017.8

TABLE I. CLASSIFICATION PERFORMANCES WITH SMALL-SCALE TARGET DATABASES

	Target database modality		
	X-ray	MRI	CT
Direct transfer learning using VGG16	81.5%	44.7%	85.8%
Modality-bridge transfer learning with bridge data of the same acquisition modality	90.1%	71.4%	91.4%
Transfer learning with bridge data of the different acquisition modality	66.4%	55.3%	80.4%

In our experiment, VGG16 network [25] was employed as the projection and classification function, which is trained by ImageNet dataset. There were a total of 16 layers, which are 13 convolutional layers, 2 fully-connected layers, and a softmax layer. The 13 convolutional layers and 2 fully connected layers were considered as the projection function.

Deep Transfer learning



Methods	10FCV	Evaluation		
	DS_Original	DS_Original	DS_Aug1	DS_Aug2
Baseline_2015 [10,14]	-	-	60.91	-
CNN-6	59.95	57.09	58.33	66.13
VGGNet-16	87.27	67.83	70.50	71.61
ResNet-50	89.34	72.10	75.75	76.78
Our proposed model	90.22	72.42	76.07	76.87

Table 2. Accuracy of visual methods in ImageCLEF2016.

Methods	10FCV	Evaluation		
	DS_Original	DS_Original	DS_Aug1	DS_Aug2
Baseline_2016 [23]	-	-	85.38	-
CNN-6	75.87	70.67	74.70	81.86
VGGNet-16	85.13	78.99	81.73	83.54
ResNet-50	87.47	82.51	85.25	86.92
Our proposed model	88.40	82.61	86.07	87.37

CNNs	Epochs				Learning Rate	Batch Size
	10FCV	DS_Original	DS_Aug1	DS_Aug2		
CNN-6	25	25	25	5	0.001	32
VGGNet-16	15	15	15	5	0.0002	32
ResNet-50	30	30	30	5	0.0002	32

Yuhai Yu., et al., Deep Transfer Learning for Modality Classification of Medical Images, Information 2017, 8, 91

Ensemble method for Medical image Classification

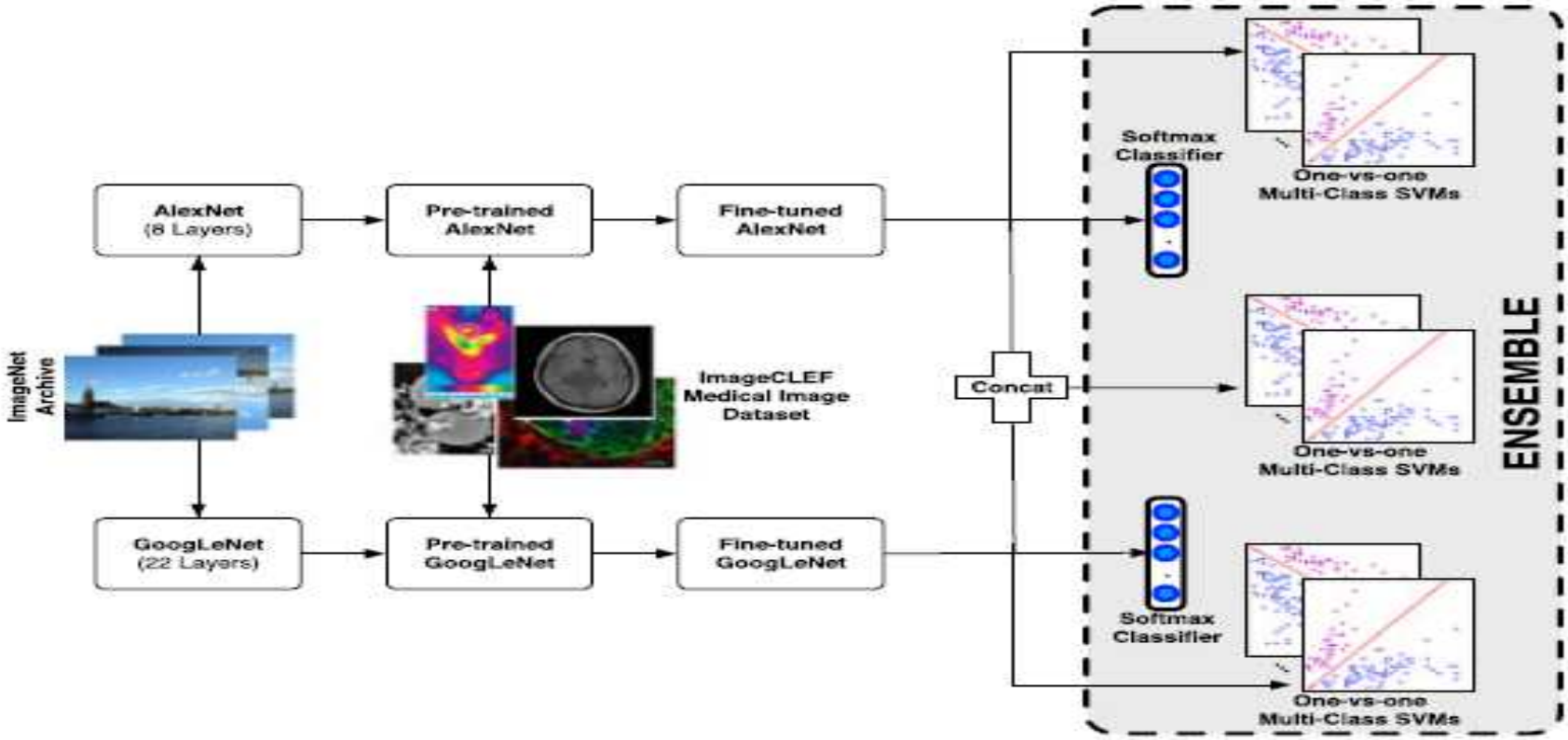
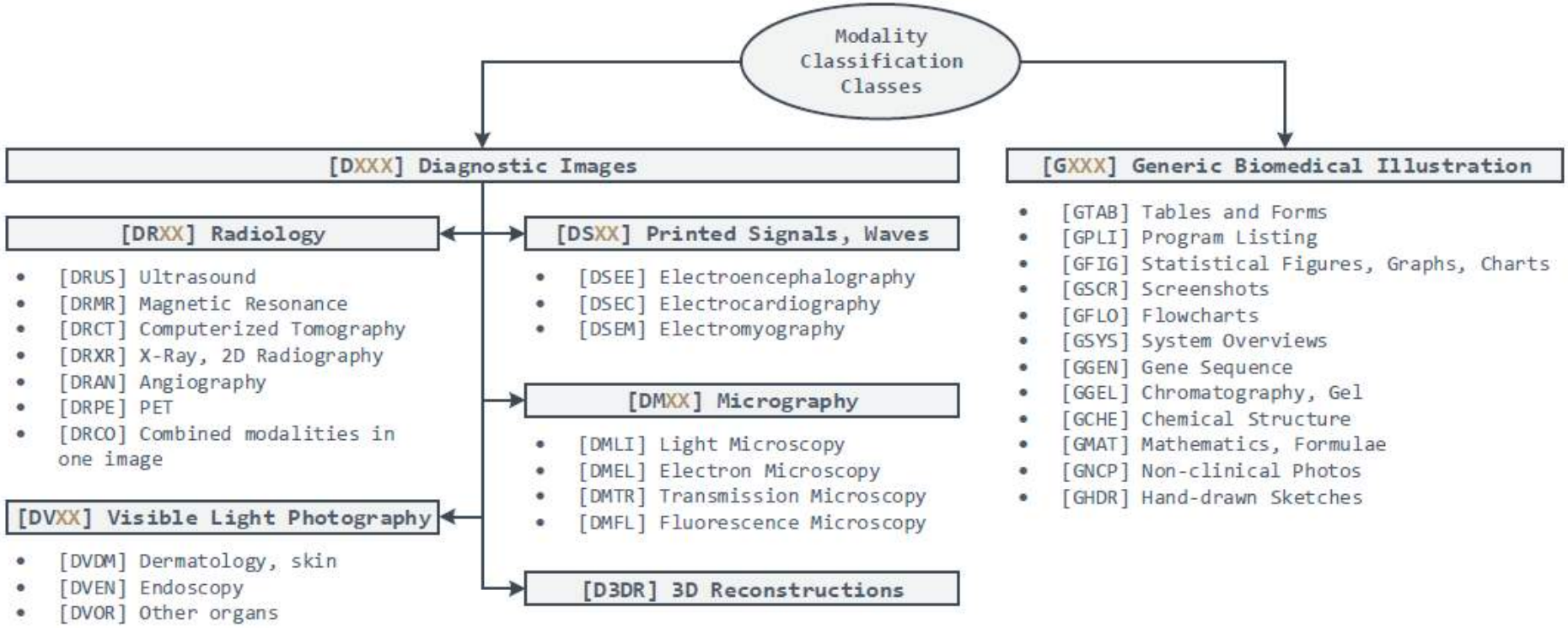


Fig. 1. Overview of our ensemble method.

Modality Classification Classes



Garca Seco de Herrera, et. al., Overview of the ImageCLEF 2015 Medical Classication task. In: Working Notes of CLEF 2015 (Cross Language Evaluation Forum). CEUR Workshop Proceedings, vol. 1391 (September 2015)

Ensemble Design

Our ensemble comprised of the following classifiers:

1) Fine-tuned AlexNet using a softmax classifier.
2) Fine-tuned GoogLeNet using a softmax classifier.
3) A one-vs-one multiclass SVM trained using features extracted from the fine-tuned AlexNet. We extracted 4096 features using the activations of the last fully connected layer of the fine-tuned network. For efficient classifier training, we reduced the dimensionality using principle component analysis (PCA) [46]. Our feature vectors were the principle components that explained 90% of the variation in the data (dimensionality: 459).
4) A one-vs-one multiclass SVM trained using features extracted from the fine-tuned GoogLeNet. We extracted 1024 features using the activations of the last pooling layer of the fine-tuned network. As above, we used PCA so that the features were the components that explained 90% of the data variation (dimensionality: 108).
5) A one-vs-one multiclass SVM trained using features extracted from the fine-tuned AlexNet and GoogLeNet. The features captured from each CNN were concatenated to form a single 5120-dimensional vector and then reduced to the principle components that explained 90% of the data variation (dimensionality: 508).

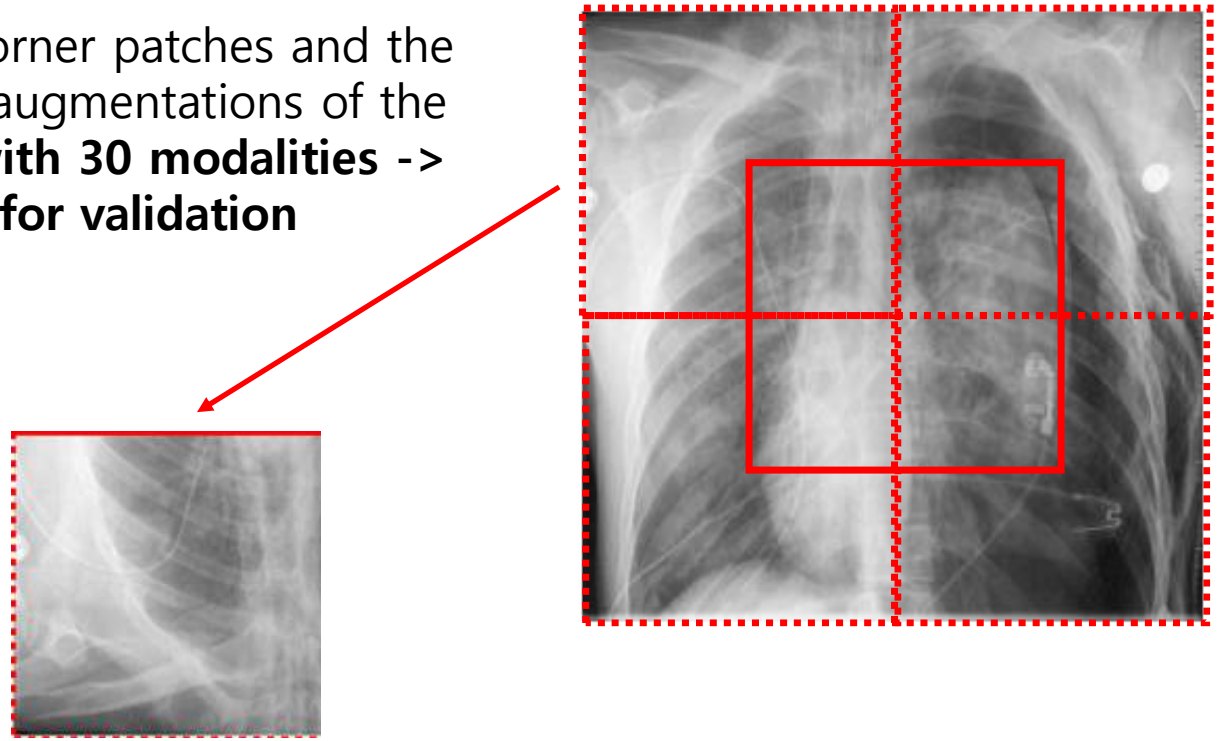
We then obtained the posterior probability $P_{i,k}(m)$ that the ith crop of the test image depicted a particular modality m according to the kth classifier in the ensemble. We determined the modality $m^\star$ of an image by fusing the posterior probabilities according to

$$m^\star = \arg\max_m \frac{\sum_i^A \sum_k^C P_{i,k}(m)}{A \times C} \tag{3}$$

where $A = 10$ is the number of augmented variations of each test image and $C = 5$ is the number of ensemble classifiers.

Data Augmentation

Image augmented by the four corner patches and the center patch +flipped image(10 augmentations of the image), **6776-ImageCLEF2016 with 30 modalities -> 67,760, 90% for training, 10% for validation**



CNN Fine-Tuning & Parameter

Let $\mathbf{X}$ be the training dataset of n images. Fine-tuning is an iterative process that finds the filter weights $\mathbf{w}$ that minimises the CNN's empirical loss (i.e., reduces the error rate)

$$L(\mathbf{w}, \mathbf{X}) = \frac{1}{n} \sum_{i=1}^{n} l\left(f\left(\mathbf{x}_i, \mathbf{w}\right), \hat{c}_i\right) \tag{1}$$

where $\mathbf{x}_i$ is the ith image of $\mathbf{X}$, $f\left(\mathbf{x}_i, \mathbf{w}\right)$ is the CNN function that predicts the class c_i of $\mathbf{x}_i$ given $\mathbf{w}$, $\hat{c}_i$ is the ground-truth class of the ith image, and $l\left(c_i, \hat{c}_i\right)$ is a penalty function for predicting c_i instead of $\hat{c}_i$. We set l to the logistic loss function.

Let $\mathbf{B} \subset \mathbf{X}$ be a subset of b images;

$b = 256, \quad \eta = 5 \times 10^{-6} \quad \alpha = 0.9 \quad \lambda = 1.$

12 GB NVIDIA Titan X GPU.

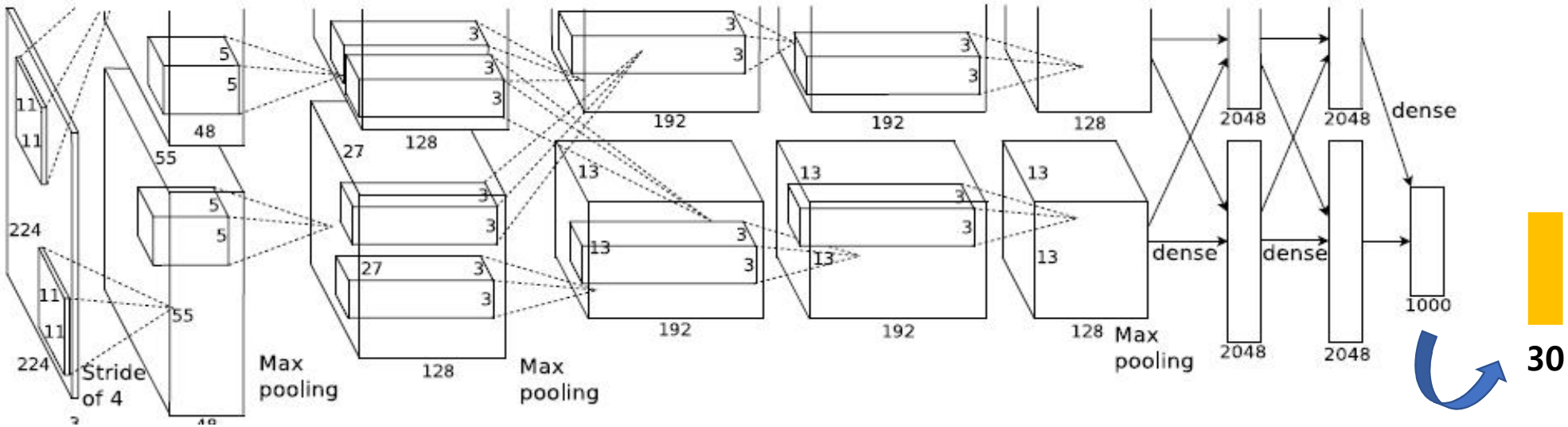
Generally speaking, one epoch consists of $\frac{n}{b}$ mini-batches of size b. However, when b is not a factor of n, the last mini-batch may have less than b images.

We iterated over the mini-batches of each epoch; the CNN weights were updated each iteration. The updated weights $\mathbf{w}_{t+1}$ were calculated from the gradient of the loss L when applied to the mini-batch $\mathbf{B}$ using the current weights $\mathbf{w}_t$

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \eta \left[\alpha \Delta \mathbf{w}_t - \frac{\partial L\left(\mathbf{w}_t, \mathbf{B}\right)}{\partial \mathbf{w}_t} - \lambda \mathbf{w}_t\right] \tag{2}$$

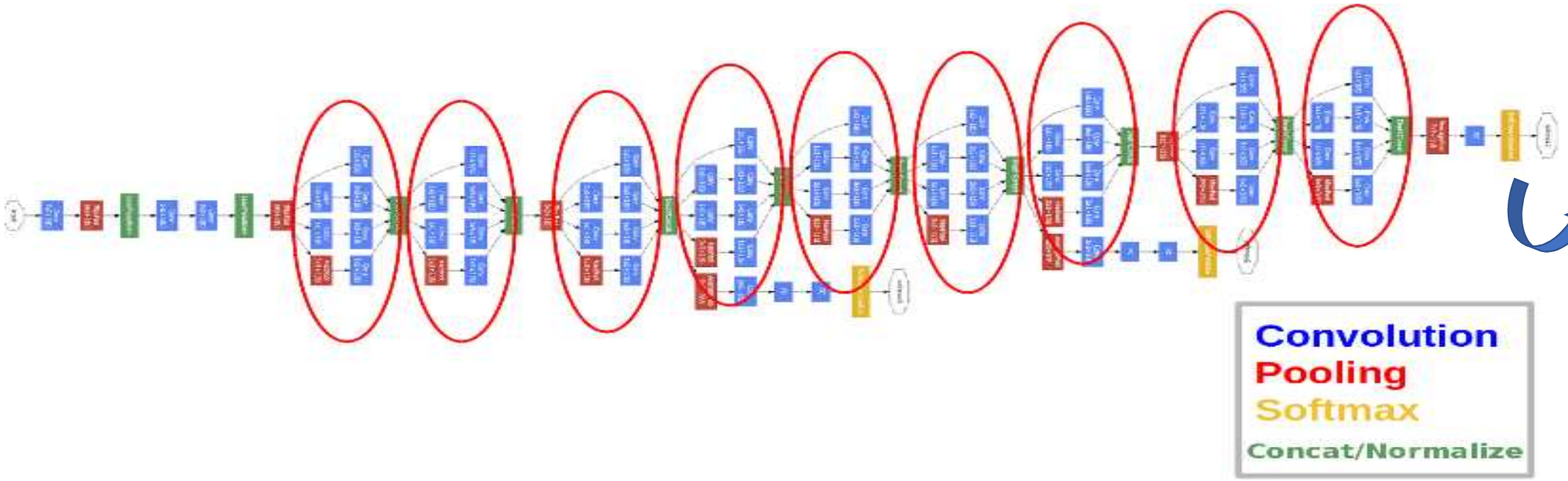
where $\Delta \mathbf{w}_t = \mathbf{w}_t - \mathbf{w}_{t-1}$ is the weight update from the previous iteration. The coefficient η is the learning rate controlling the size of the updates to the weights. The momentum coefficient α diminishes fluctuations in weight changes over consecutive iterations by adding a proportion of the previous update to the current update; this has the effect of speeding up the learning process while simultaneously smoothing the weight updates. The weight decay λ shrinks the weights to find the smallest optimal weights.

AlexNet Model



A. Krozhevsky, et al., Imagenet classification with deep convolutional neural networks, 2012 PANI.

GoogleNet Model



inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Table 1: GoogLeNet incarnation of the Inception architecture

C.Szegedy, et al., Going deeper with convolutions, IEEE Conf. Pattern Recognition, 2015

Heat Map of Confusion Matrix

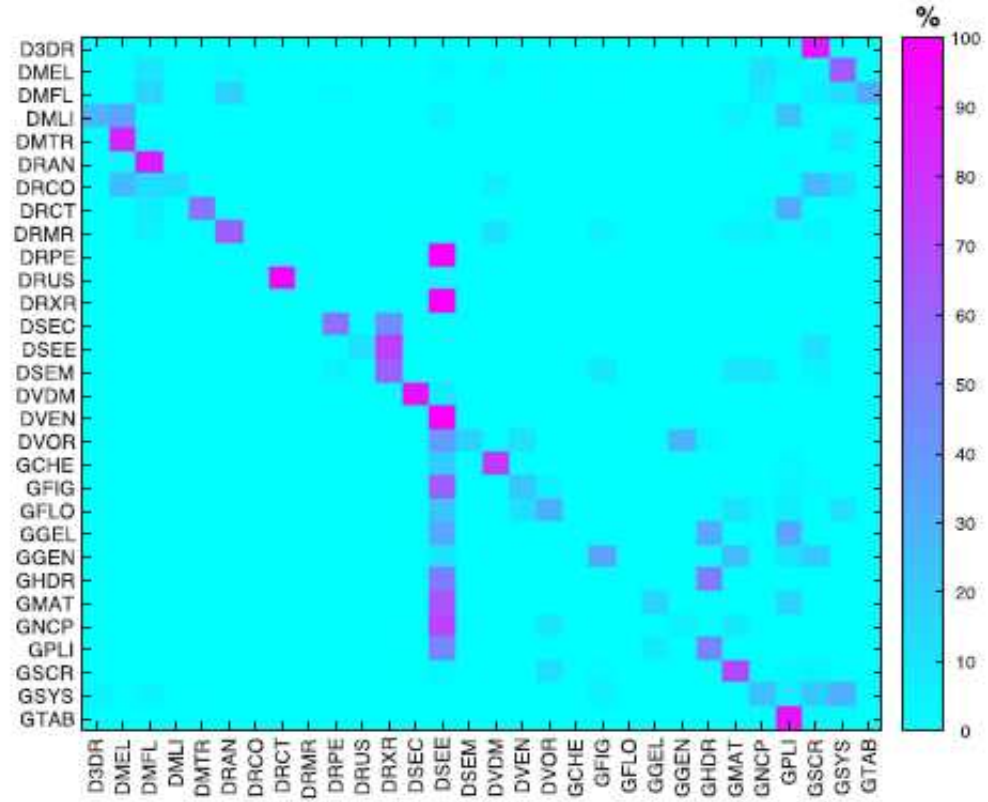


Fig. 4. Confusion matrix for our ensemble. The matrix entries have been scaled to a percentage to account for the uneven distribution of the classes.

Results

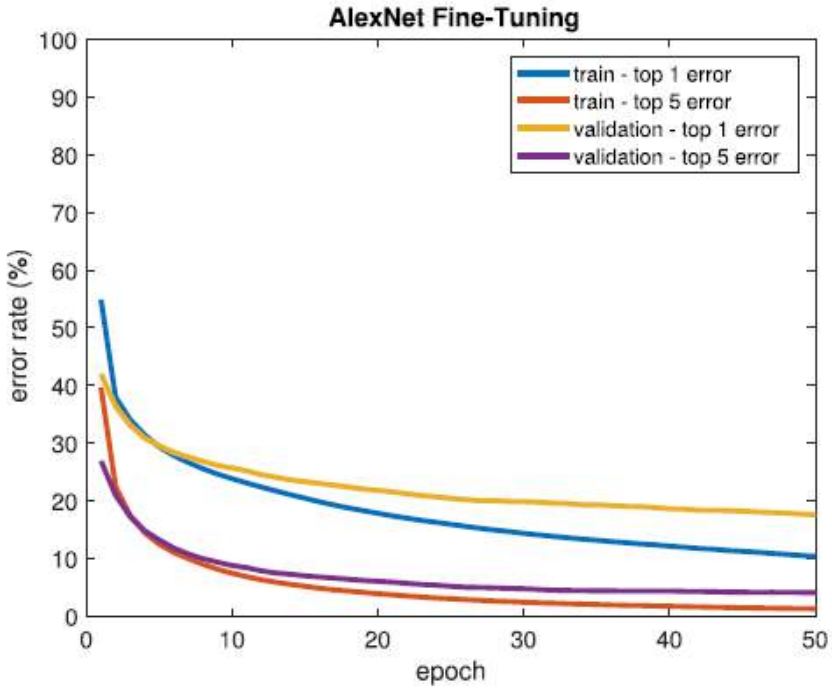


Fig. 2. Fine-tuning AlexNet over 50 epochs.

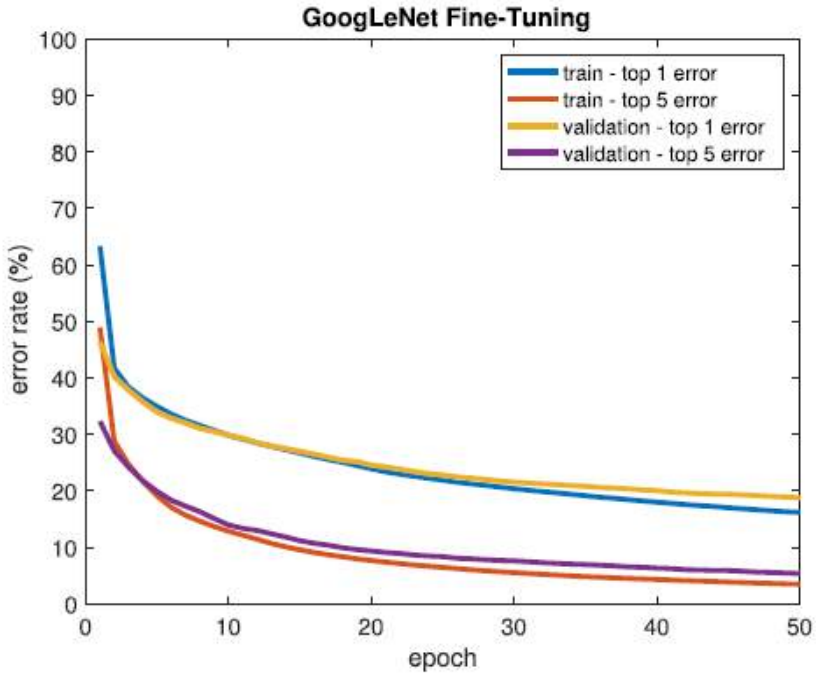


Fig. 3. Fine-tuning GoogLeNet over 50 epochs.

Results

TABLE II
TOP 1 CLASSIFICATION ACCURACY (%)

	Architecture	
	AlexNet	GoogLeNet
transfer learned + SVM	79.21	78.61
fine-tuned + softmax	79.62	77.17
fine-tuned + SVM	79.60	80.75
our ensemble	82.48	

TABLE III
TOP 5 CLASSIFICATION ACCURACY (%)

	Architecture	
	AlexNet	GoogLeNet
transfer learned + SVM	96.71	96.33
fine-tuned + softmax	94.48	91.31
fine-tuned + SVM	96.47	96.54
our ensemble	96.59	

TABLE IV
COUNT OF IMAGES CORRECTLY CLASSIFIED BY A PAIR OF METHODS

		AlexNet[a]			GoogLeNet[a]			
		TLS	FTC	FTS	TLS	FTC	FTS	ENS
Alex	**TLS**		3087	3113	3047	3000	3102	**3188**
	FTC	3087		3118	3024	3036	3107	**3231**
	FTS	3113	3118		3038	2991	3086	**3211**
GoogLe	**TLS**	3047	3024	3038		2974	3085	**3142**
	FTC	3000	3036	2991	2974		3065	**3132**
	FTS	3102	3107	3086	3085	3065		**3235**
	ENS	**3188**	**3231**	**3211**	**3142**	**3132**	**3235**	

[a] TLS = transfer learned + SVM, FTC = fine-tuned + CNN softmax, FTS = fine-tuned + SVM, ENS = our ensemble

Results

TABLE V
TOP 1 ACCURACY COMPARED TO OTHER METHODS

Method	Accuracy (%)
transfer learned ResNet-152 [37]*	85.38
hand-crafted feature collection [37]*	84.46
RGB color PHOW [20]*	84.01
our ensemble	*82.48*
RGB color PHOW [20]	81.73
modified GoogLeNet (60 epochs) [37]*	81.03
fine-tuned AlexNet (100 epochs) [36]	77.55
VGG-like CNN (500 epochs) [38]	65.31

[*] training dataset expanded with additional examples

Results

Group	Class	# Samples		Results			
		Train	Test	Precision	Sensitivity	Specificity	F-Score
3-D reconstructions (D3DR)		201	96	69.31	72.92	99.24	71.07
microscopy	electron microscopy (DMEL)	208	88	39.62	23.86	99.22	29.79
	fluorescence microscopy (DMFL)	906	284	73.45	91.55	97.58	81.50
	light microscopy (DMLI)	696	405	87.94	91.85	98.64	89.86
	transmission microscopy (DMTR)	300	96	48.39	62.50	98.43	54.55
radiology	angiography (DRAN)	17	76	92.31	31.58	99.95	47.06
	combined modalities (DRCO)	33	17	41.67	29.41	99.83	34.48
	computerised tomography (DRCT)	61	71	80.26	85.92	99.63	82.99
	magnetic resonance (DRMR)	139	144	75.29	90.97	98.93	82.39
	positron emission tomography (DRPE)	14	15	100	13.33	100	23.53
	ultrasound (DRUS)	26	129	98.59	54.26	99.98	70.00
	x-ray, 2-D radiography (DRXR)	51	18	34.38	61.11	99.49	44.00
signals	electrocardiography (DSEC)	10	8	0	0	100	0
	electroencephalography (DSEE)	8	3	100	100	100	100
	electromyography (DSEM)	5	6	0	0	100	0
photos	dermatology, skin (DVDM)	29	9	62.50	55.56	99.93	58.82
	endoscopy (DVEN)	16	8	100	12.50	100	22.22
	other images (DVOR)	55	21	52.00	61.90	99.71	56.52
generic biomedical illustrations	chemical structure (GCHE)	61	14	92.86	92.86	99.98	92.86
	statistics, figures, graphs, charts (GFIG)	2954	2085	88.80	99.23	87.46	93.73
	flowcharts (GFLO)	20	31	71.43	16.13	99.95	26.32
	chromatography, gel (GGEL)	344	224	95.03	76.79	99.77	84.94
	gene sequence (GGEN)	179	150	71.74	22.00	99.68	33.67
	hand-drawn sketches (GHDR)	136	49	29.09	32.65	99.05	30.77
	mathematics, formula (GMAT)	15	3	0	0	100	0
	non-clinical photos (GNCP)	88	20	36.84	35.00	99.71	35.90
	program listing (GPLI)	1	2	0	0	100	0
	screenshots (GSCR)	33	6	50.00	16.67	99.98	25.00
	system overviews (GSYS)	91	75	33.33	6.67	99.76	11.11
	tables and forms (GTAB)	79	13	50.00	46.15	99.86	48.00

Discussion & Conclusion

- Sallower networks as AlexNet -> more generalizable. features that are applicable to a wider variety of images.
- Deeper networks as GoogleNet -> more semantically meaningful features.
- Ensemble -> Combined Sallower + Deeper.
- One-vs-one Multiclass SVM -> subtle difference compared to the Classifier of CNN Softmax.
- Ensemble -> Distinguish between image modalities with subtle differences.