

# Sentiment Analysis Update

Quanzeng You 2014/05/19

VISTA @ URCS

# Performance on testing data

Testing_boost	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC
model4	12970	3243	3331	12011	0.8	0.8	0.8	0.78	0.8	0.79	0.85
resume	13815	2398	3078	12264	0.82	0.85	0.85	0.8	0.83	0.83	0.86
resume_400000	12952	3261	3408	11934	0.79	0.80	0.80	0.78	0.80	0.79	0.84
boost_training	12654	3559	3905	11437	0.76	0.78	0.78	0.75	0.77	0.76	0.82
model2	11830	4383	2857	12485	0.81	0.73	0.73	0.81	0.77	0.77	0.84
Testing	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC
model4	20964	7788	8378	20175	0.71	0.73	0.73	0.71	0.72	0.72	0.79
resume	23760	4992	7540	21013	0.76	0.83	0.83	0.74	0.79	0.78	0.86
resume_400000	21084	7668	9077	19476	0.70	0.73	0.73	0.68	0.72	0.71	0.78
boost_training	20429	8323	10073	18480	0.67	0.71	0.71	0.65	0.69	0.68	0.74
model2	19723	9029	6177	22376	0.76	0.69	0.69	0.78	0.72	0.73	0.80

# Twitter dataset

Select 1270 images from 1.5 million images.

Need more? Good for now. Start AMT

# Baseline

- ★ SentiBank
  - 1200 classifiers
  - SVM for 10-fold cross-validation
  - Accuracy: 78%
    - All candidate images have been classified as positive
    - May need to use our Twitter for cross-validation
- ★ Other Baselines
  - Low-level and mid-level

# Sentiment Analysis Update

Quanzeng You 2014/05/19

VISTA @ URCS

# Sentiment Analysis Update

Quanzeng You 2014/05/12

VISTA @ URCS

# Performance on testing data

Testing_boost	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC
model4	12970	3243	3331	12011	0.8	0.8	0.8	0.78	0.8	0.79	0.85
resume	13815	2398	3078	12264	0.82	0.85	0.85	0.8	0.83	0.83	0.9
resume_400000	12952	3261	3408	11934	0.79	0.80	0.80	0.78	0.80	0.79	0.84
boost_training	12654	3559	3905	11437	0.76	0.78	0.78	0.75	0.77	0.76	0.82
Testing	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC
model4	20964	7788	8378	20175	0.71	0.73	0.73	0.71	0.72	0.72	0.79
resume	23760	4992	7540	21013	0.76	0.83	0.83	0.74	0.79	0.78	0.86
resume_400000	21084	7668	9077	19476	0.70	0.73	0.73	0.68	0.72	0.71	0.78
boost_training	20429	8323	10073	18480	0.67	0.71	0.71	0.65	0.69	0.68	0.74

# Performance on training data

Training_Boost	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC
model4	187771	3254	4537	174266	0.98	0.98	0.98	0.97	0.98	0.98	1
resume	182581	8444	16552	162251	0.92	0.96	0.96	0.91	0.94	0.93	0.98
resume_400000	187120	3905	5818	172985	0.97	0.98	0.98	0.97	0.97	0.97	0.99
boost_training	188322	2703	4898	173905	0.97	0.99	0.99	0.97	0.98	0.98	1.00
Training	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC
model4	244673	13028	18355	239693	0.93	0.95	0.95	0.93	0.94	0.94	0.98
resume	223775	33926	54439	203609	0.8	0.87	0.87	0.79	0.84	0.83	0.91
resume_400000	237422	20279	30354	227694	0.89	0.92	0.92	0.88	0.90	0.90	0.97
boost_training	217454	40247	52212	205836	0.81	0.84	0.84	0.80	0.82	0.82	0.88

# Performance on Twitter

	Twitter_VS O											
	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC	
model4	365	78	86	37	0.81	0.82	0.82	0.30	0.82	0.71	0.57	
resume	364	79	83	40	0.81	0.82	0.82	0.33	0.82	0.71	0.62	
	Labeled by me											
	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC	
model4	194	30	36	15	0.84	0.87	0.87	0.29	0.85	0.76	0.64	
resume	199	25	34	17	0.85	0.89	0.89	0.33	0.87	0.79	0.69	

# Download more Twitter Images

-New 410650 images

Now, a total of 376 images.

# Sentiment Analysis Update

Quanzeng You 2014/05/05

VISTA @ URCS

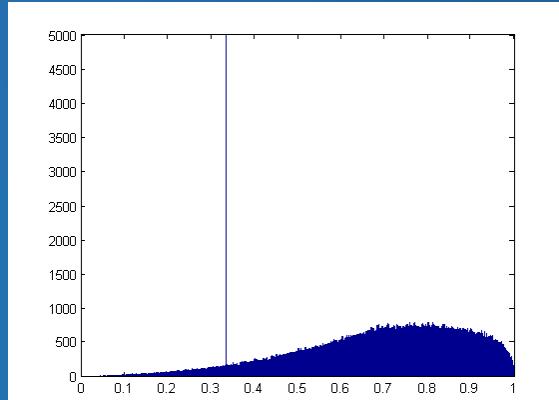
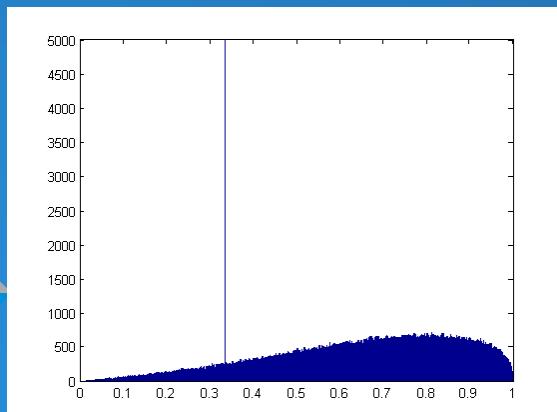
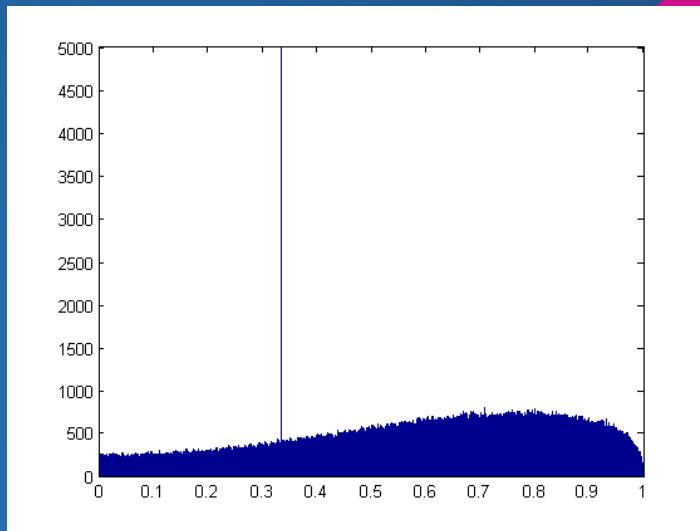
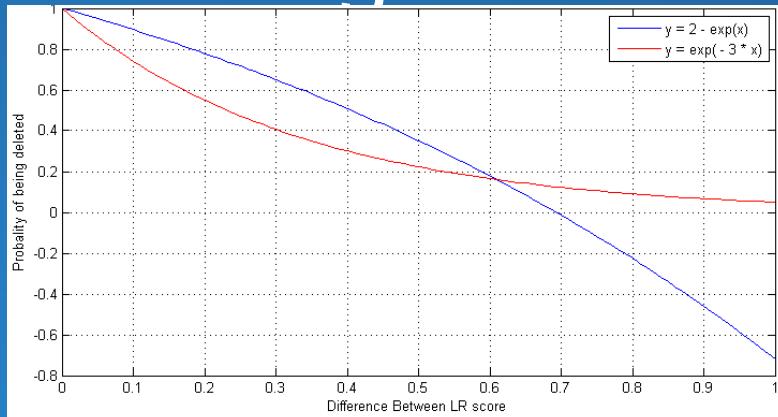
# Create balanced data set

- ★ For each negative image
  - With prob ( |pos| - |neg| ) / |neg| duplicate a center crop from resized image (300 for the smaller edge)

(Start with Boosting is still running)

	negative	positive	total		Boosting		
training	258048	257701	515749		178803	191025	369828
testing	28553	28752	57305		15342	16213	31555

# Boosting



# Performance on testing data

	Testing_boost											
	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC	
model4	12970	3243	3331	12011	0.80	0.80	0.80	0.78	0.80	0.79	0.85	
resume	13815	2398	3078	12264	0.82	0.85	0.85	0.80	0.83	0.83	0.90	
	Testing											
	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC	
model4	20964	7788	8378	20175	0.71	0.73	0.73	0.71	0.72	0.72	0.79	
resume	23760	4992	7540	21013	0.76	0.83	0.83	0.74	0.79	0.78	0.86	

# Performance on training data

	Training											
	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC	
model4	244673	13028	18355	239693	0.93	0.95	0.95	0.93	0.94	0.94	0.98	
resume	223775	33926	54439	203609	0.80	0.87	0.87	0.79	0.84	0.83	0.91	
	Training_boost											
	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC	
model4	187771	3254	4537	174266	0.98	0.98	0.98	0.97	0.98	0.98	1.00	
resume	182581	8444	16552	162251	0.92	0.96	0.96	0.91	0.94	0.93	0.98	

# Performance on Twitter

	Twitter_VS O											
	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC	
model4	365	78	86	37	0.81	0.82	0.82	0.30	0.82	0.71	0.57	
resume	364	79	83	40	0.81	0.82	0.82	0.33	0.82	0.71	0.62	
	Labeled by me											
	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC	
model4	194	30	36	15	0.84	0.87	0.87	0.29	0.85	0.76	0.64	
resume	199	25	34	17	0.85	0.89	0.89	0.33	0.87	0.79	0.69	

# Sentiment Analysis Update

Quanzeng You 2014/04/28

VISTA @ URCS

# Configuration

90% Training data(401739)

10% Testing data (44637)

## ★ After sampling

- Training data(374065)
- Testing data(40462)

Other parameters are the same with previously model.

# Performance on Testing(44637)

	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC
model4	24412	4141	10824	5260	0.693	0.855	0.855	0.327	0.765	0.665	0.660
model4-boost	25079	3474	11332	4752	0.689	0.878	0.878	0.295	0.772	0.668	0.659
Boost-resume	25028	3525	11342	4742	0.688	0.877	0.877	0.295	0.771	0.667	0.658

# Performance on Boost Testing(40462)

	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC
model4	23044	3243	9786	4389	0.702	0.877	0.877	0.310	0.780	0.678	0.662
model4-boost	23408	2879	10121	4054	0.698	0.890	0.890	0.286	0.783	0.679	0.662
Boost-resume	23372	2915	10092	4083	0.698	0.889	0.889	0.288	0.782	0.679	0.660

# Performance on Twitter

	tp	fn	fp	tn	precision	recall	Sensitivity	Specificity	F1	accuracy	AUC
model4	399	44	101	22	0.80	0.90	0.90	0.18	0.85	0.74	0.61
boost	408	35	109	14	0.79	0.92	0.92	0.11	0.85	0.75	0.59
boost-resume	413	30	108	15	0.79	0.93	0.93	0.12	0.86	0.76	0.62

# Sentiment Analysis Update

Quanzeng You 2014/04/21

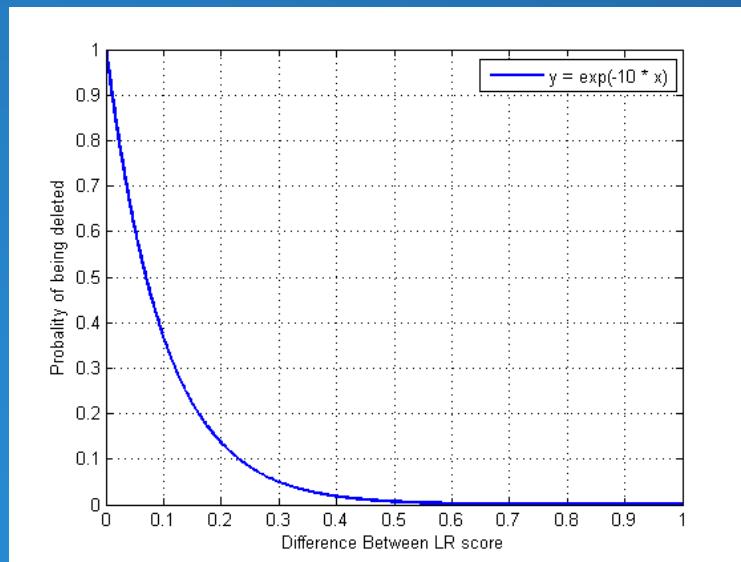
VISTA @ URCS

# Bug Fix

- ★ Mean file for python
  - Using a python wrapper for prediction
    - Numpy.npy format
    - Different from Caffenet (Google Protocol Buffer)
  - Wrong order of R and B channel.
    - All results need to be recalculated.

# Progressive CNN

- ★ Probabilistically select samples
  - Related to the difference of LR score



# New Dataset

- ★ Total 404610
  - Training: 384380
  - Testing: 20230
- ★ Old whole data set 446376
  - 424058
  - 22318

# Training

- ★ Start with the new data set to train (Boost)
- ★ Using the new data set to resume the training of the old network. (Boost-Resume)
  - From the saved 200000-th state to resume the training using the new data set for another 100000 iterations.

# Performance on the Old Dataset

Total training instances: 424058

Total testing instances: 22318

	Training	Testing	Training New	Testing New
Model4	0.8837847653	0.6648445201	0.87880482	0.8805734058
Boost	0.8369609818	0.8434447531	NA	0.6871972318
Boost-Resume	0.8971461451	0.7656600054	0.9114287944	0.8432525952

# Performance on New Dataset

Still running

Total training: 384380

Total testing: 20230

	Training	Testing
Model4	0.878804828555	0.8805734058
Boost	NA	0.6871972318
Boost-Resume	0.9114287944	0.8432525952

# Performance on Twitter

Model4	0.7632508834
Boost	0.7491166078
Boost-Resume	0.7597173145

# Sentiment Analysis Update

Quanzeng You 2014/04/14

VISTA @ URCS

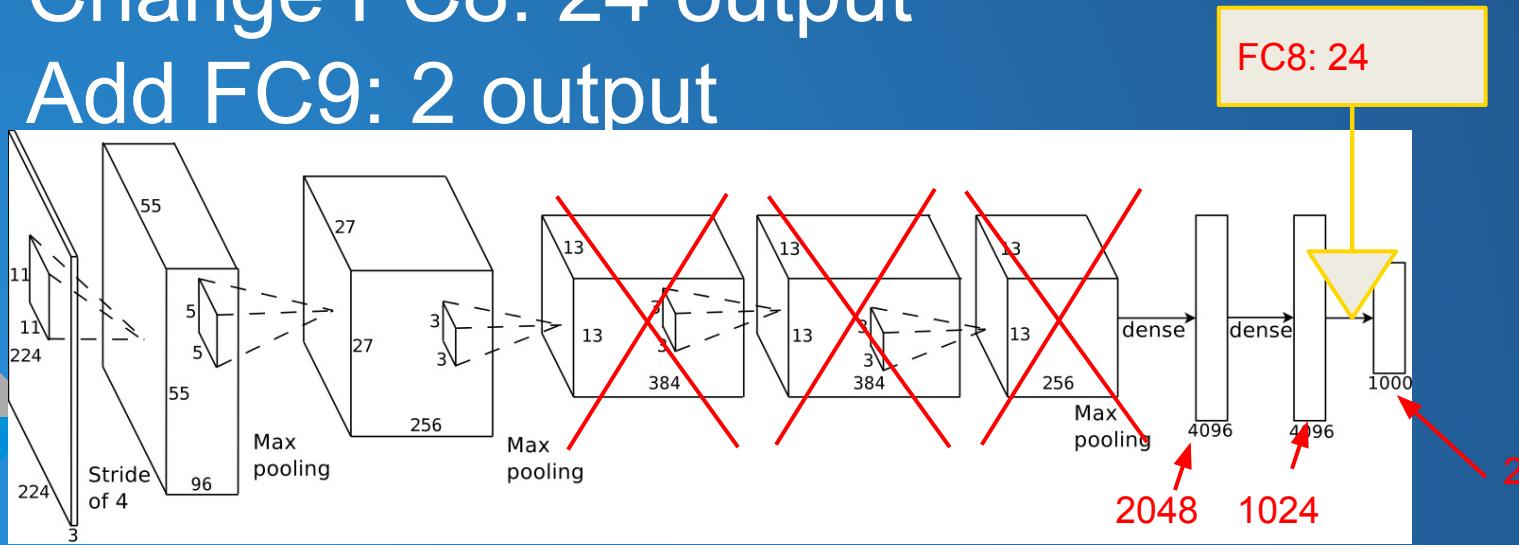
# Lower-Level Model 5

Delete one conv-layer

Reduce FC6 and FC7 4096  $\rightarrow$  2048, 1024

Change FC8: 24 output

Add FC9: 2 output



# Performance of model5

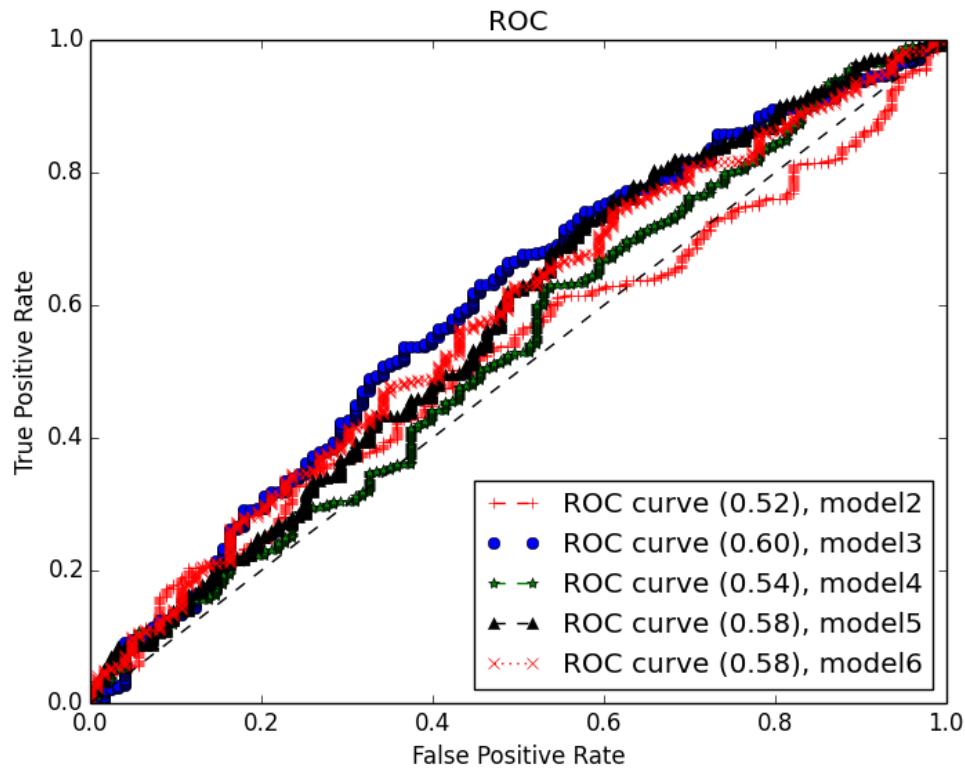
- ★ Twitter
  - Accuracy is 0.749116607774
- ★ Training
  - Accuracy is 0.843460092723
- ★ Testing
  - Accuracy is 0.657137736356

# Ensemble for Twitter

Total of 566 samples

	tp	fn	fp	tn	precision	recall	F1	accuracy	AUC
Model2	388	55	113	10	0.7745	0.8758	0.8220	0.7032	0.5246
Model3	409	34	107	16	0.7926	0.9233	0.8530	0.7509	0.5969
Model4	423	20	109	14	0.7951	0.9549	0.8677	0.7721	0.5396
Model5	402	41	101	22	0.7992	0.9074	0.8499	0.7491	0.5759
Ensemble	430	13	116	7	0.7875	0.9707	0.8696	0.7721	0.5791

# Ensemble for Twitter

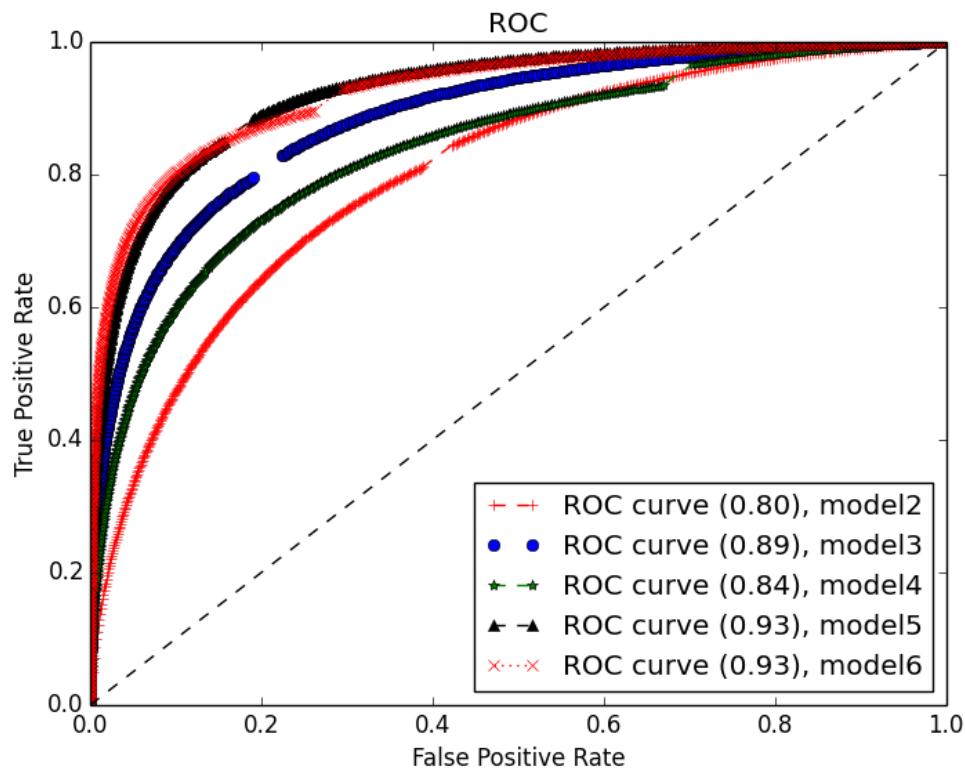


# Ensemble for Training

★ Total of 424058 samples

	tp	fn	fp	tn	precision	recall	F1	accuracy	AUC
Model2	247861	24257	85298	66642	0.7440	0.9109	0.8190	0.7417	0.8008
Model3	256960	15158	74809	77131	0.7745	0.9443	0.8510	0.7878	0.8862
Model4	248278	23840	85631	66309	0.7435	0.9124	0.8194	0.7418	0.8393
Model5	257325	14793	51589	100351	<b>0.8330</b>	0.9456	<b>0.8858</b>	<b>0.8435</b>	<b>0.9263</b>
Ensemble	263395	8723	73371	78569	0.7821	<b>0.9679</b>	0.8652	0.8064	0.9271

# Ensemble for Training

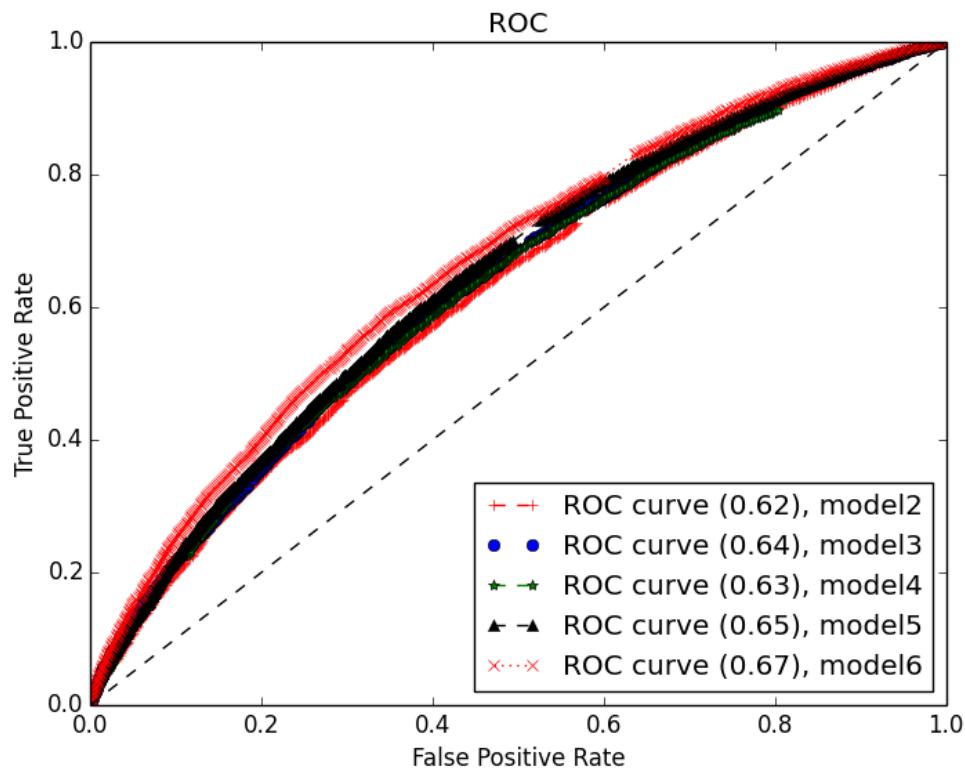


# Ensemble for Testing

★ Total of 22318 samples

	tp	fn	fp	tn	precision	recall	F1	accuracy	AUC
Model2	12121	2215	5728	2254	0.6791	0.8455	0.7532	0.6441	0.6220
Model3	12534	1802	5914	2068	0.6794	0.8743	0.7646	0.6543	0.6357
Model4	12353	1983	5865	2117	0.6781	0.8617	0.7589	0.6484	0.6320
Model5	12145	2191	5461	2521	<b>0.6898</b>	0.8472	0.7604	0.6571	0.6465
Ensemble	13034	1302	6109	1873	0.6809	<b>0.9092</b>	<b>0.7786</b>	<b>0.6679</b>	<b>0.6657</b>

# Ensemble for Testing



# Sentiment Analysis Update

Quanzeng You 2014/04/07

VISTA @ URCS

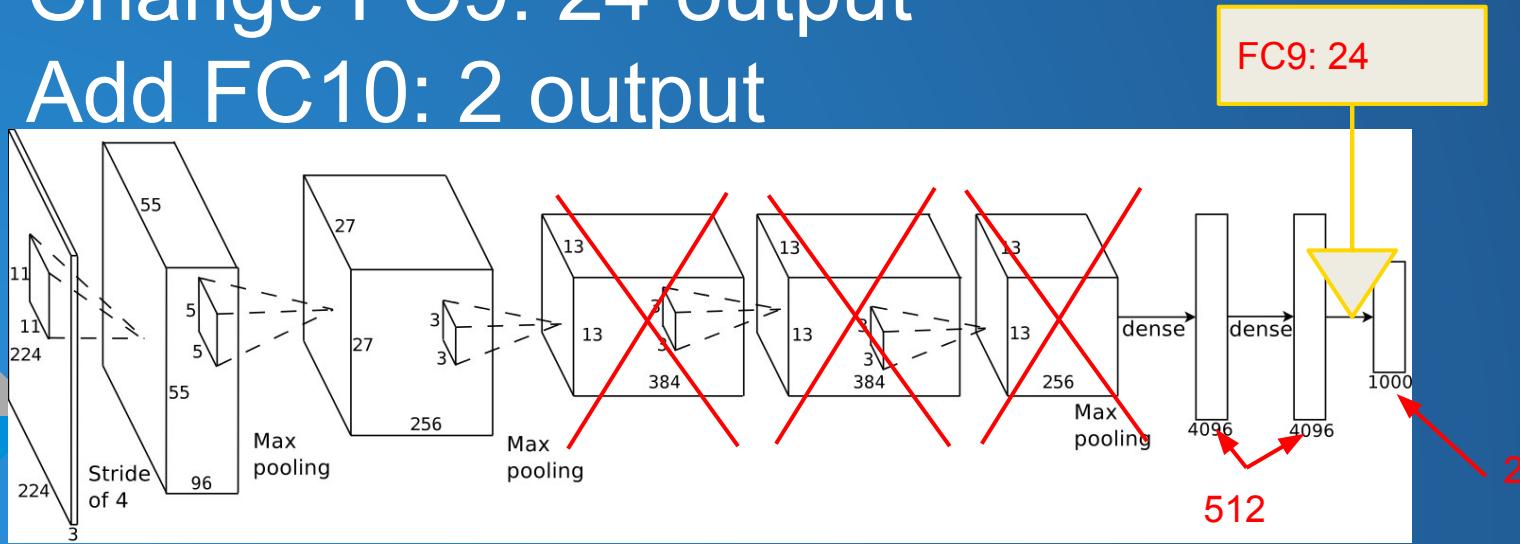
# Model 4

Delete one conv-layer

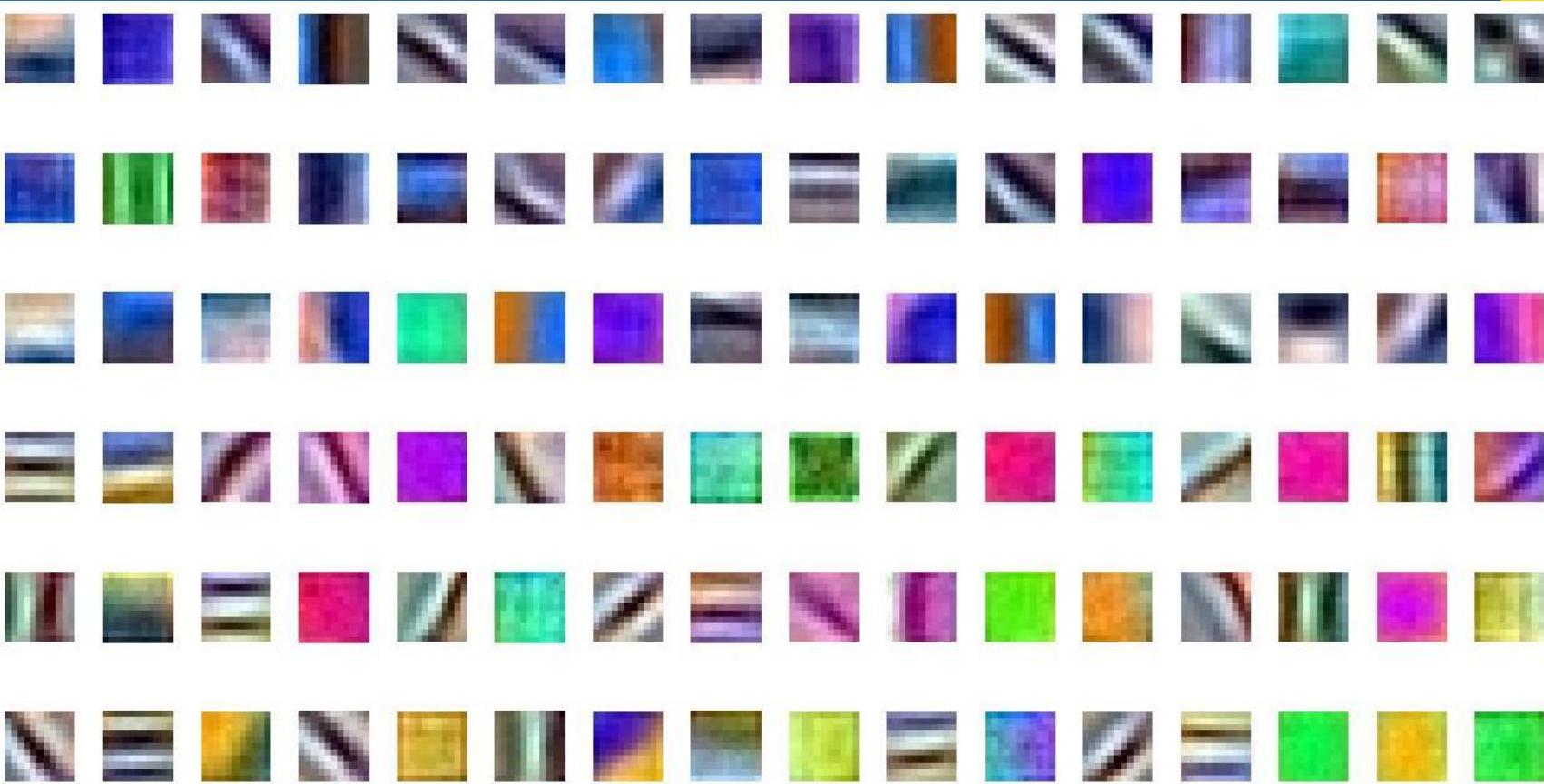
Reduce FC7 and FC8 4096  $\rightarrow$  512

Change FC9: 24 output

Add FC10: 2 output



# Weight visualization



# Accuracy of Model 4

- ★ On Testing Data (5%)
  - Total of 22318 samples, and 14470 of them are correct
  - Accuracy is 0.648355587418
- ★ On Training Data (95%)
  - Total of 424058 samples, and 314587 of them are correct
  - Accuracy is 0.741848992355

# Accuracy on Twitter Test Data

Total of 566 samples, and 437 of them are correct

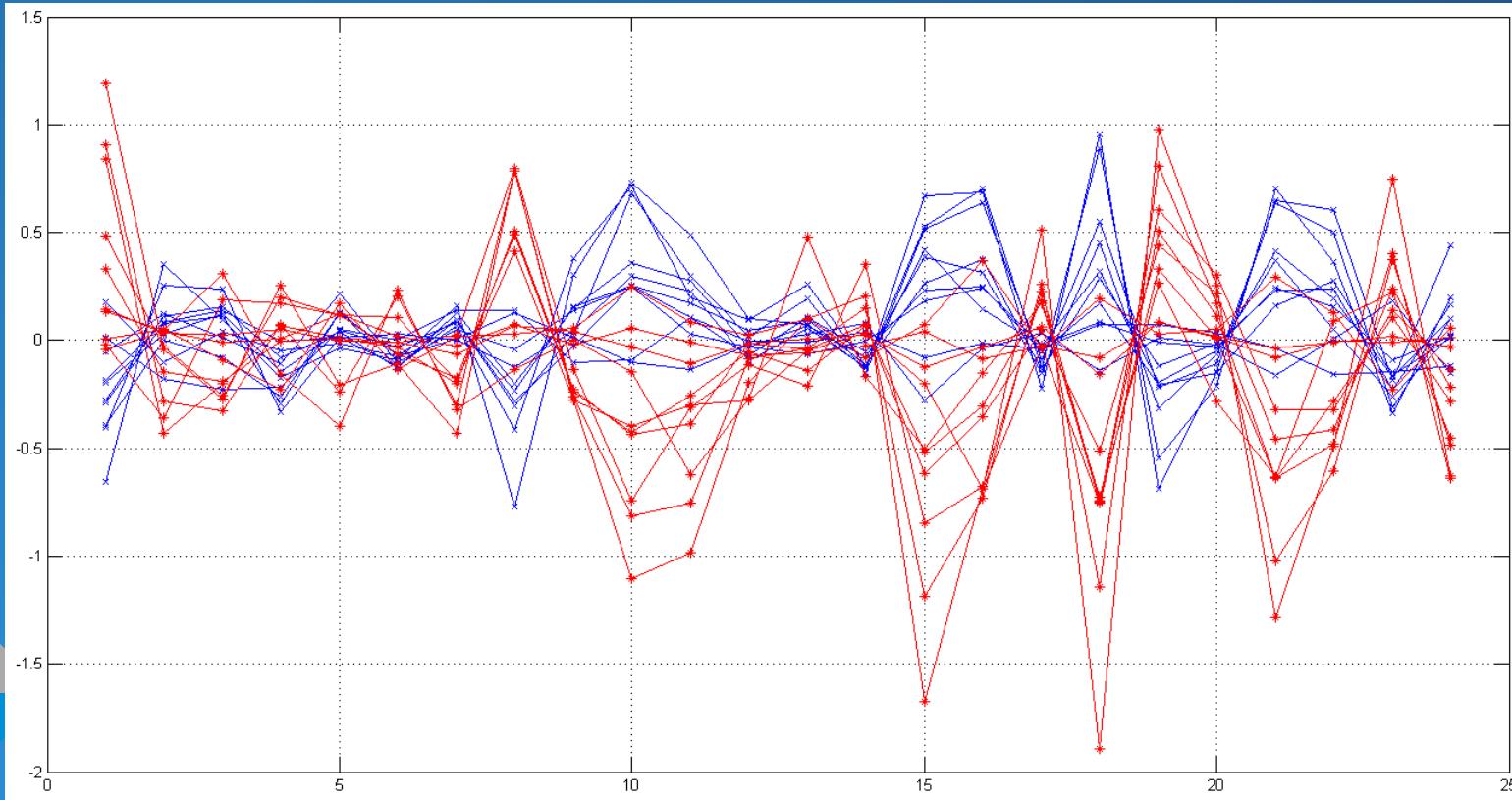
Accuracy is 0.772084805654

# Comparison

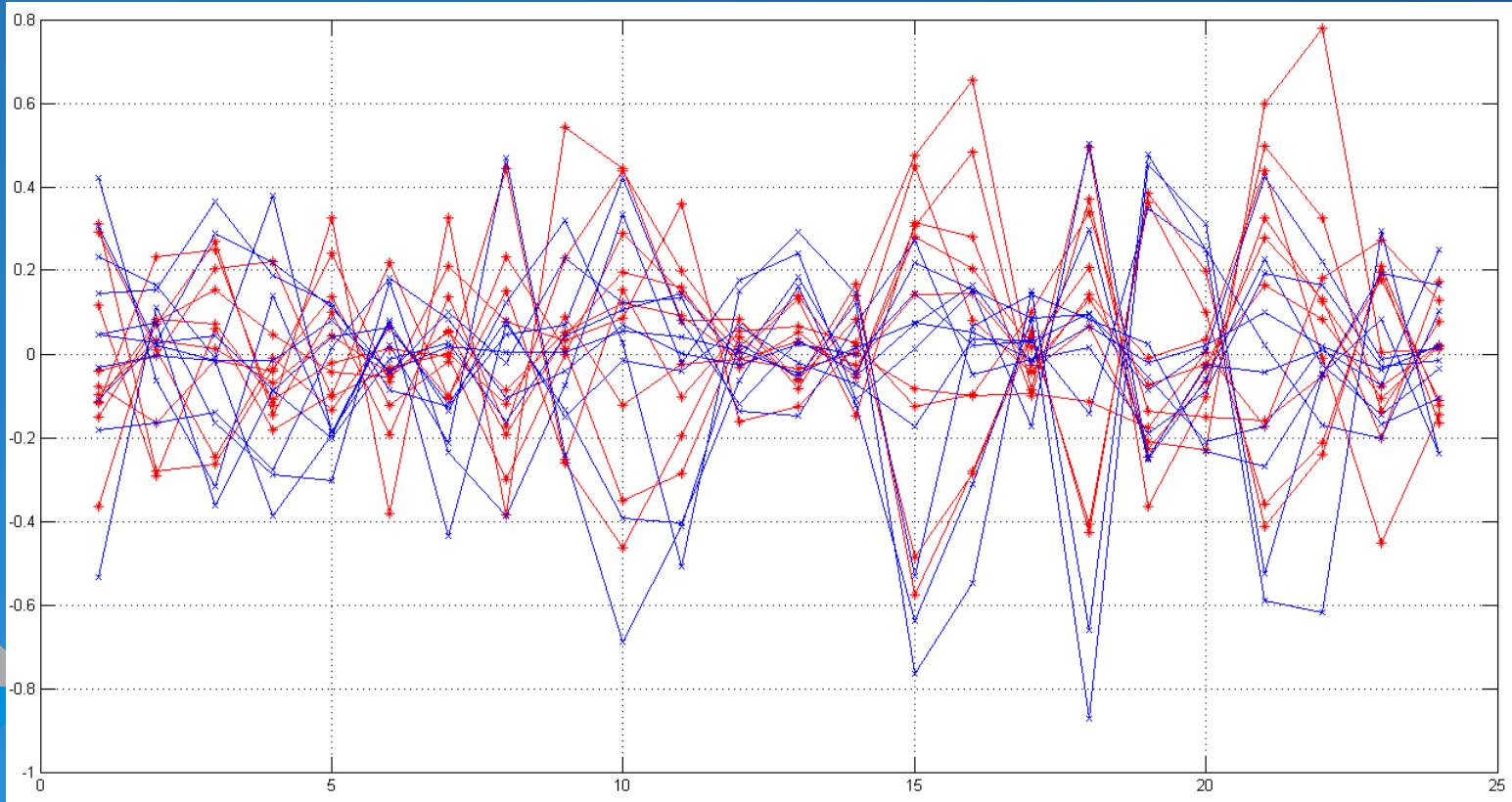
## Results on training, testing and Twitter

Model	Testing	Training	Twitter
Model2 (Deepest)	0.6441	0.7416	0.7031
Model3 (Lowest)	<b>0.6544</b>	<b>0.7811</b>	0.7509
Model4 (Middle)	0.6484	0.7418	<b>0.7720</b>

# How about the 24 outputs?



# Wrongly classified examples



# Sentiment Analysis Update

Quanzeng You

2014/03/24~2014/03/31

VISTA @ URCS

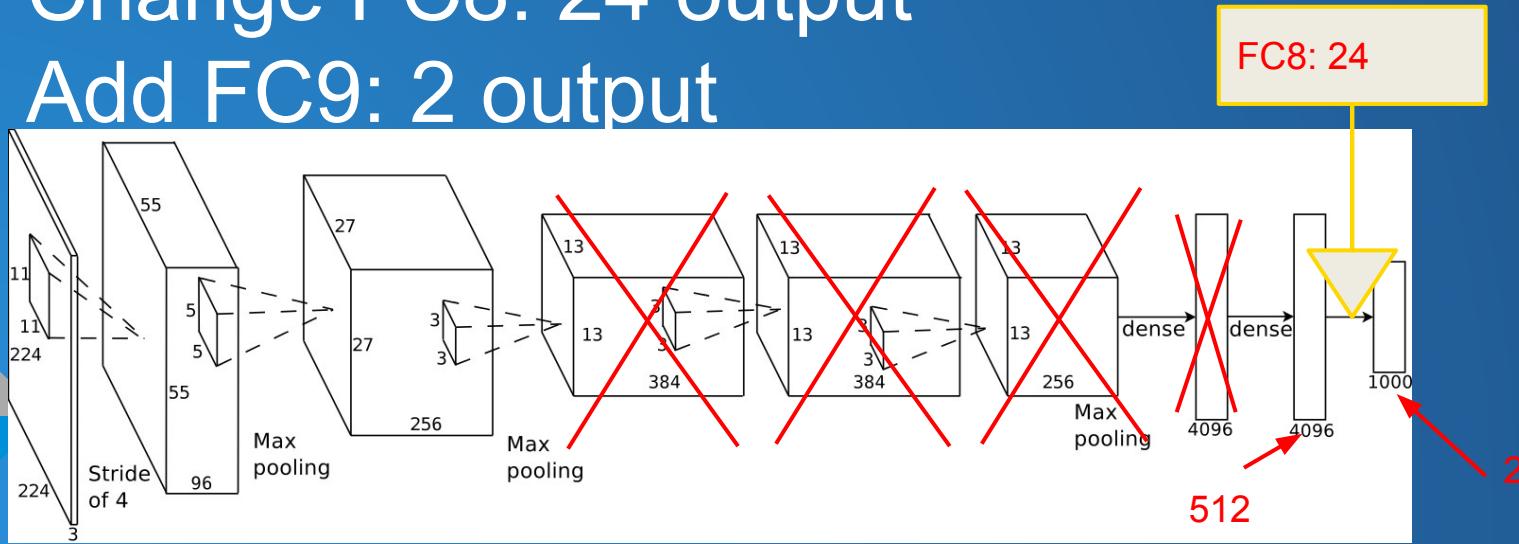
# Lower-Level Model 3

Delete one conv-layer

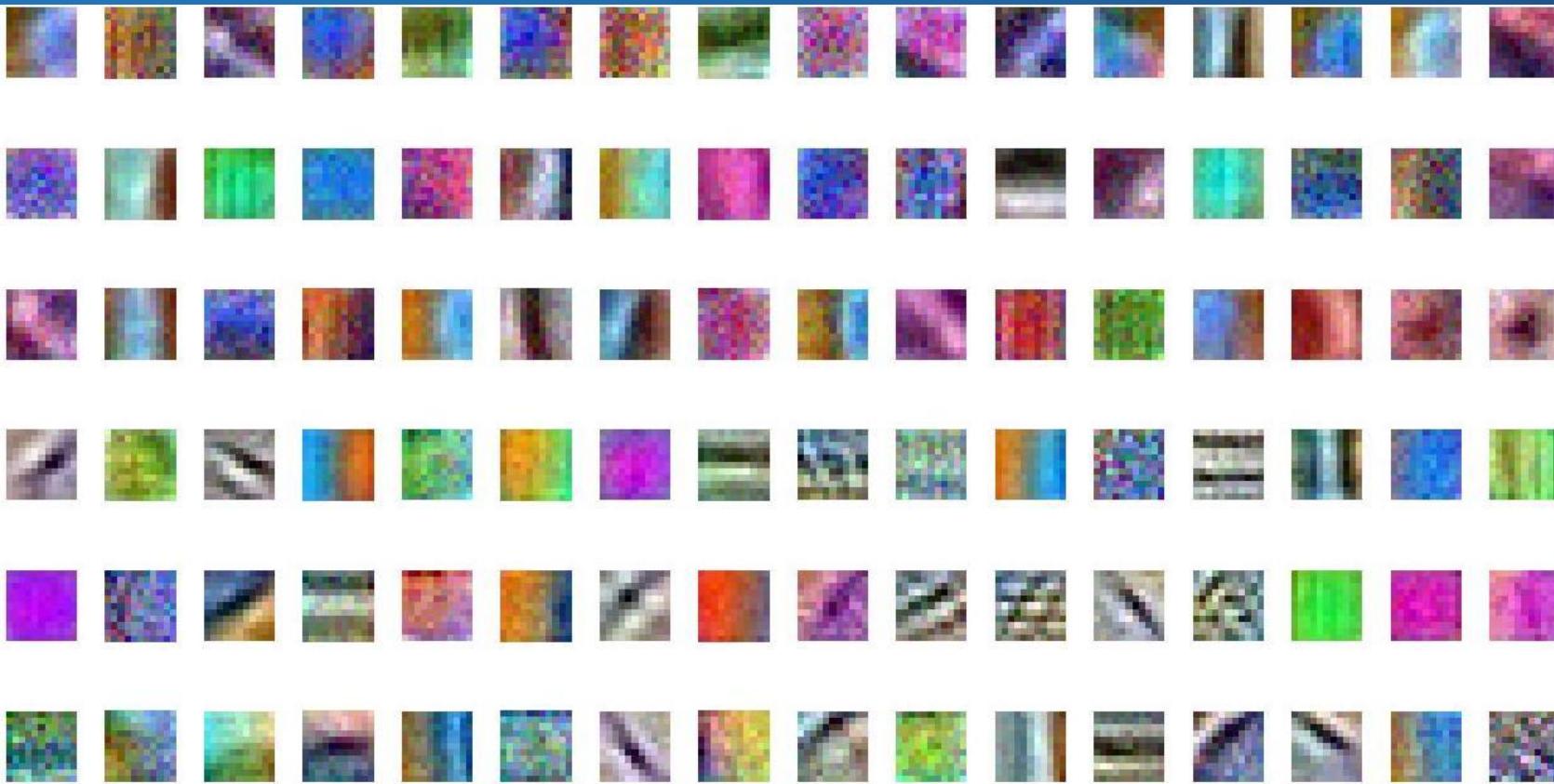
Reduce FC6 and FC7 4096  $\rightarrow$  512

Change FC8: 24 output

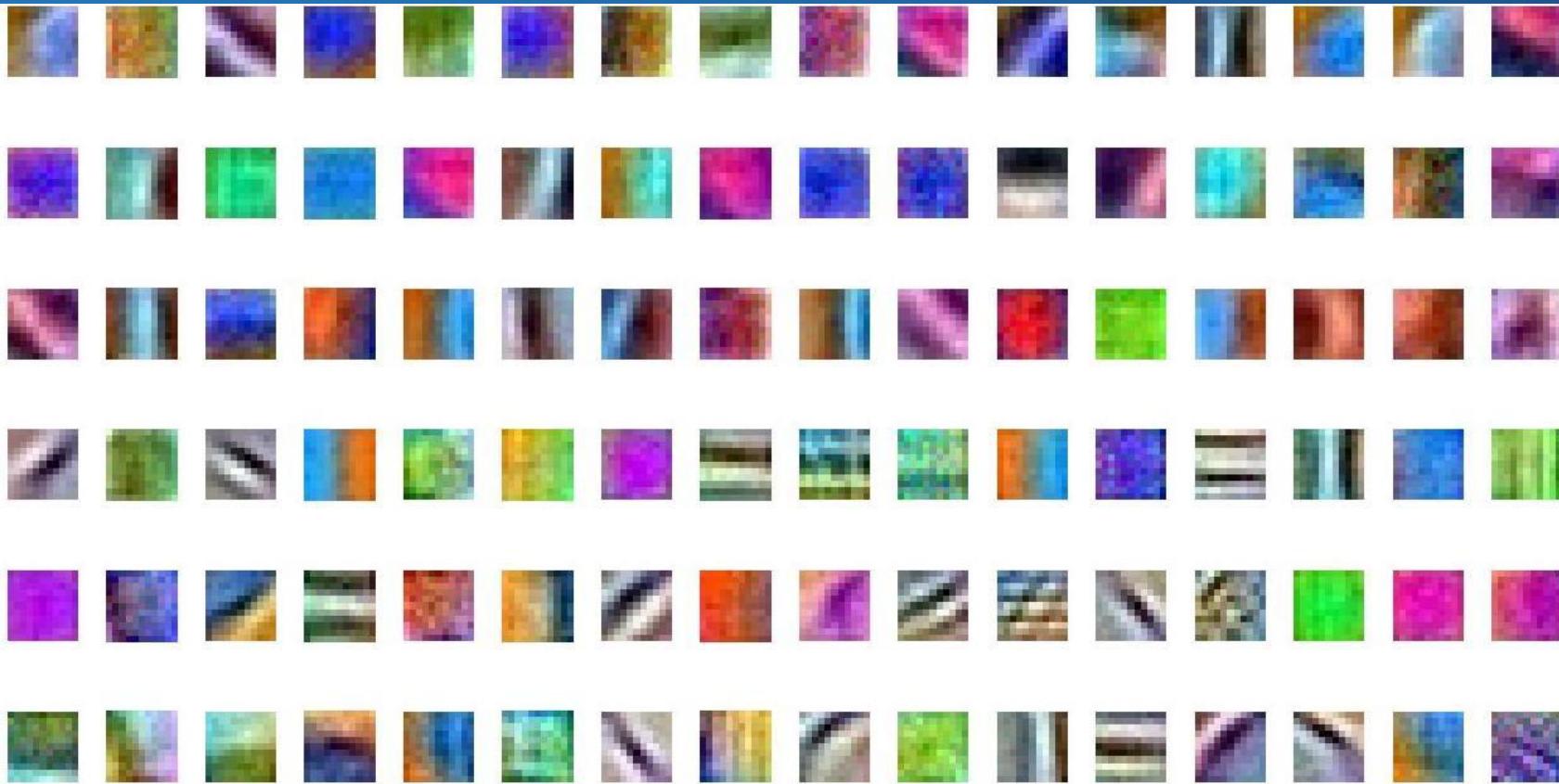
Add FC9: 2 output



# Weight filters after 300000 iterations



# Still running ( weights on the snapshot of 50000)



# Accuracy of Model 3

- ★ On Testing Data (5%)
  - Total of 22318 samples, and 14607 of them are correct
  - Accuracy is 0.654494130298
- ★ On Training Data (95%)
  - Total of 424058 samples, and 331228 of them are correct
  - Accuracy is 0.781091265817

# Accuracy on Twitter Test Data

Total of 566 samples, and 425 of them are correct

Accuracy is 0.750883392226

**Better than SentiBank!** <http://goo.gl/AfELH2>

**Table 3:** Tweet Sentiment Prediction Accuracy (Visual Based Methods)

	Linear SVM	Logistic Regr.
Low-level Features	0.55	0.57
SentiBank	0.67	<b>0.70</b>

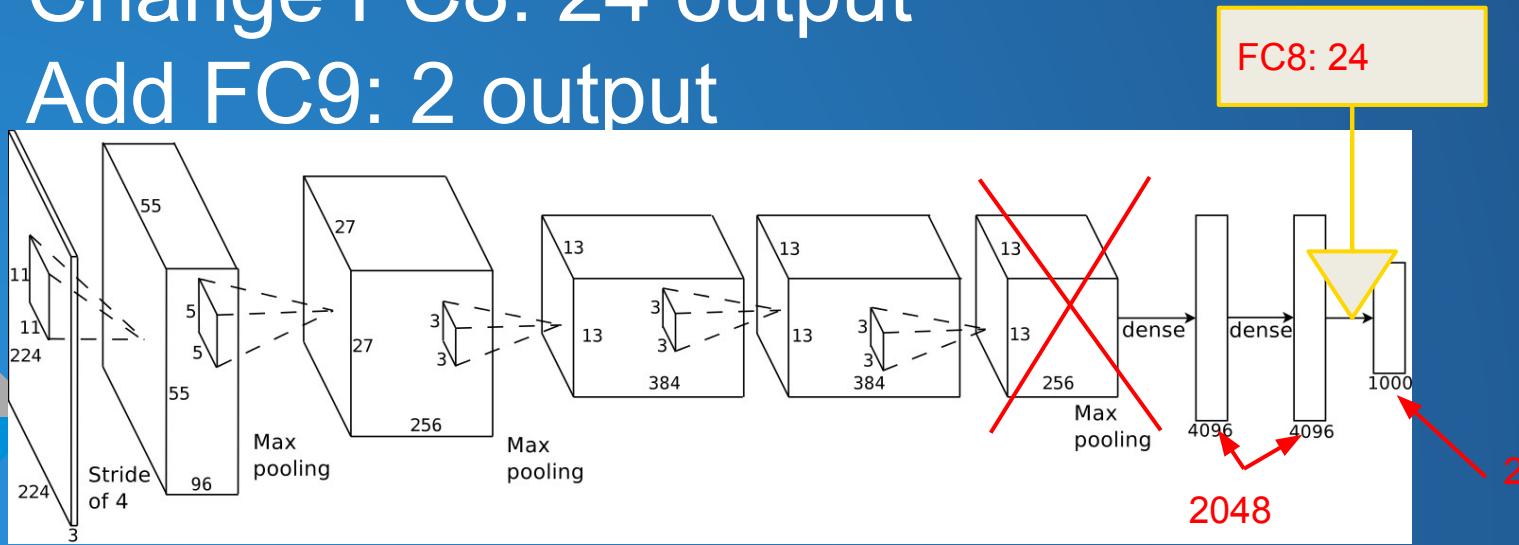
# Model 2

Delete one conv-layer

Reduce FC6 and FC7 4096  $\rightarrow$  2048

Change FC8: 24 output

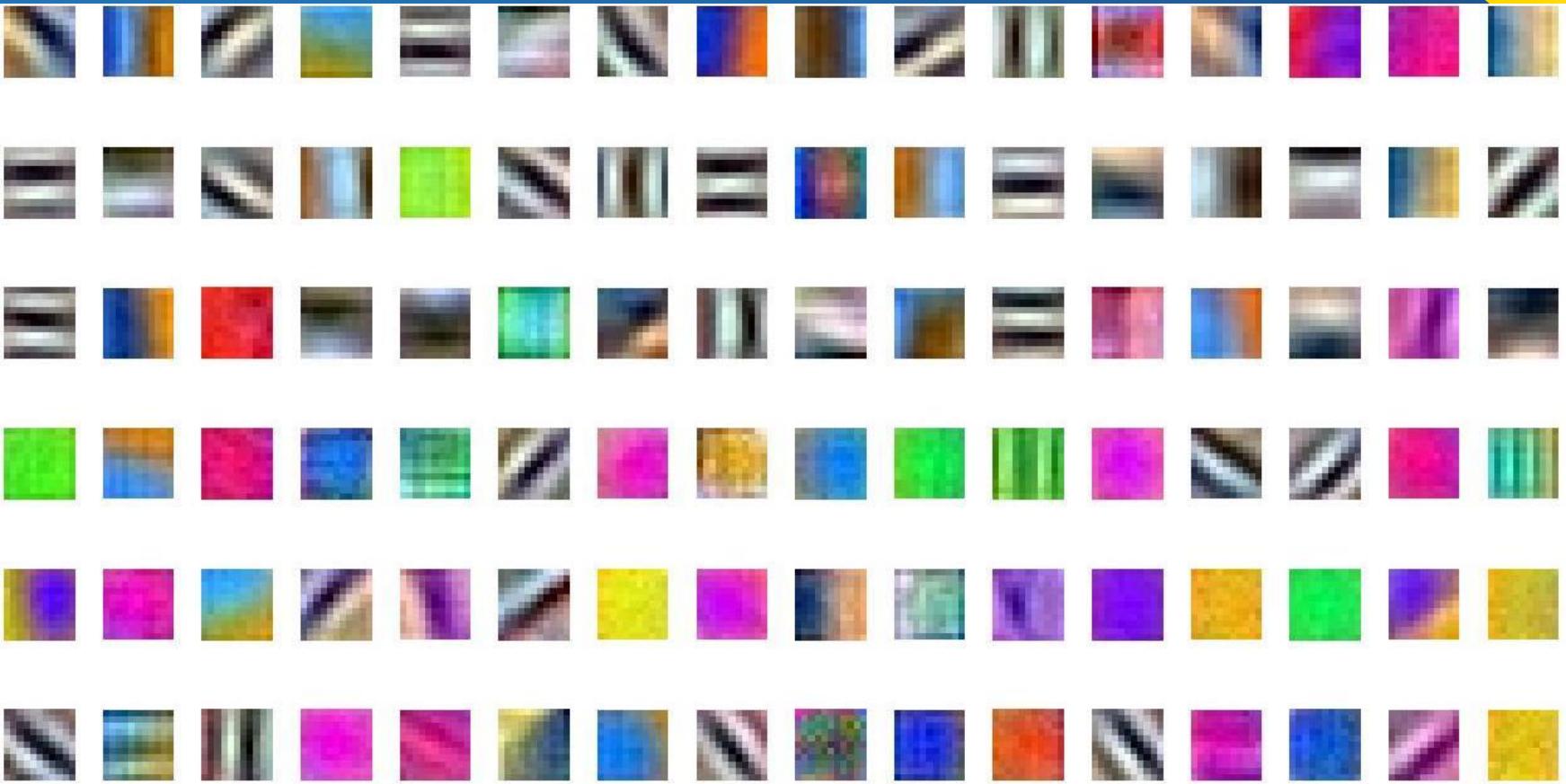
Add FC9: 2 output



# Configuration

- ★ Same data and parameters with Model 3.
- ★ More conv layers and fully connected layers
  - Slower than Model 3
    - 34 seconds vs 13 seconds (every 20 batches)
  - Still running
    - Can get results this week

# Weight filters after 300000 iterations



# Accuracy of Model 2

- ★ On Testing Data (5%)
  - Total of 22318 samples, and 14375 of them are correct
  - Accuracy is 0.644098933596
- ★ On Training Data (95%)
  - Total of 424058 samples, and 314503 of them are correct
  - Accuracy is 0.741650906244

# Accuracy on Twitter Test Data

Total of 566 samples, and 398 of them are correct

Accuracy is 0.703180212014

Worse than Model 3

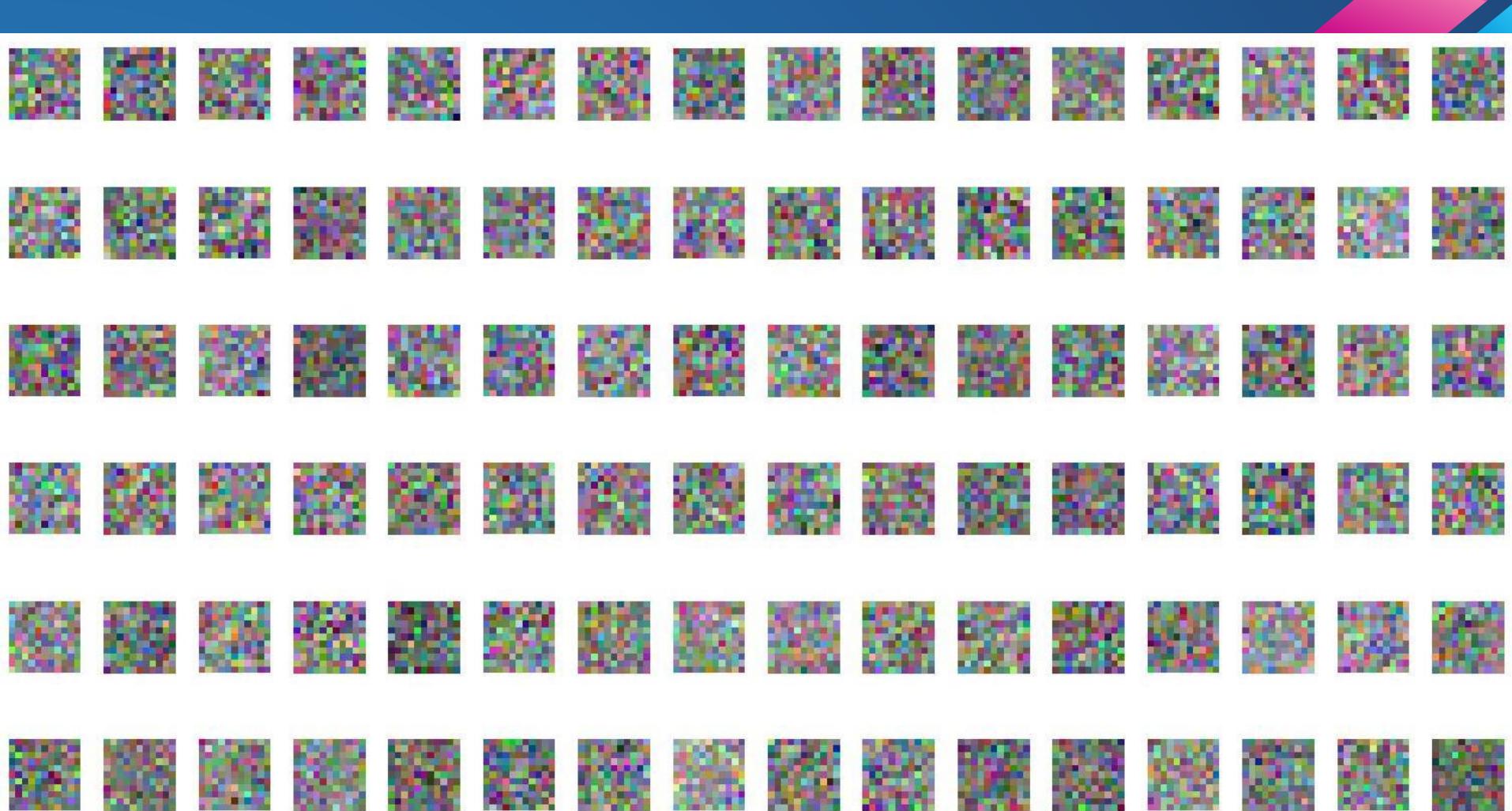
# Sentiment Analysis Update

Quanzeng You 2014/03/17

VISTA @ URCS

# Weight Visualization

- ★ Develop code to extract weight and visualize the weights from Caffe
- ★ Do experiments and fix some bugs



# How about a small data set?

I randomly choose only two ANPs.

broken\_glass and cute\_dog



# Bug on previous data set

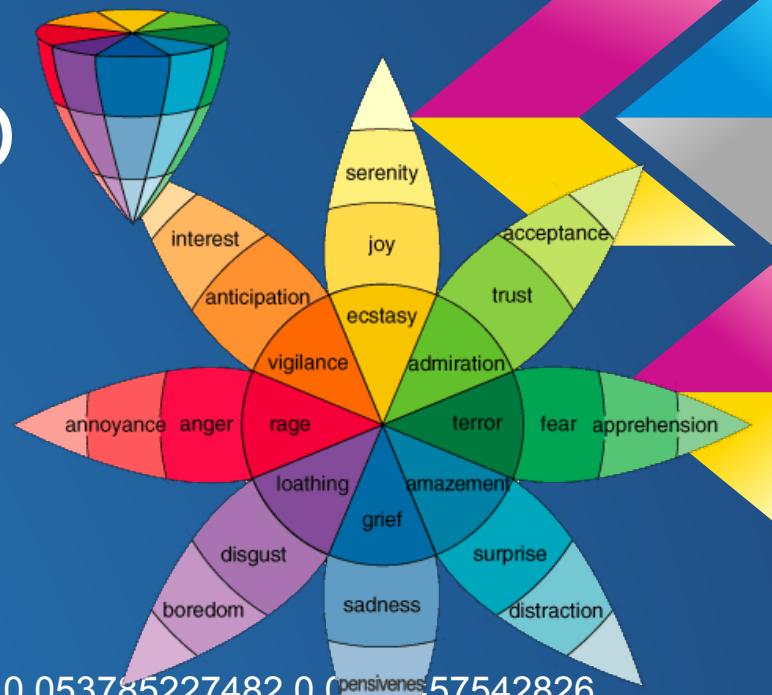
The data set is used to train classifiers for ANPs.

The fullset includes a testset of 1:1500 positive/negative ratio. Each reduced set includes a testset of 1:2 positive/negative ratio. Most results in our paper are based on the performance averaged over these 5 reduced sets. Please find the training and testing partition details in our paper.

# Additional info from VSO

- ★ Wheel of emotions (24)
- ★ 1178 pairs have emotion score

```
ill_child -1.00 1100 0.34. 0.0053416835513 0.0454964081783 0.053785227482 0.0  
0.00110517590717 0.0281819856327 0.0230244980659 0.0138146988396 0.00441777491  
0.0178670104992 0.0232086940505 0.00128937189169 0.033155277215 0.166697365997  
0.0486277399153 0.0434702523485 0.0449438202247 0.100386811568 0.0110517590717  
0.0176828145146 0.0117885430098 0.0235770860195 0.245907165224 0.0147356787622
```



# Newly Build Data Set

A total of 446377

Pos/Neg: 0.641731/0.358269

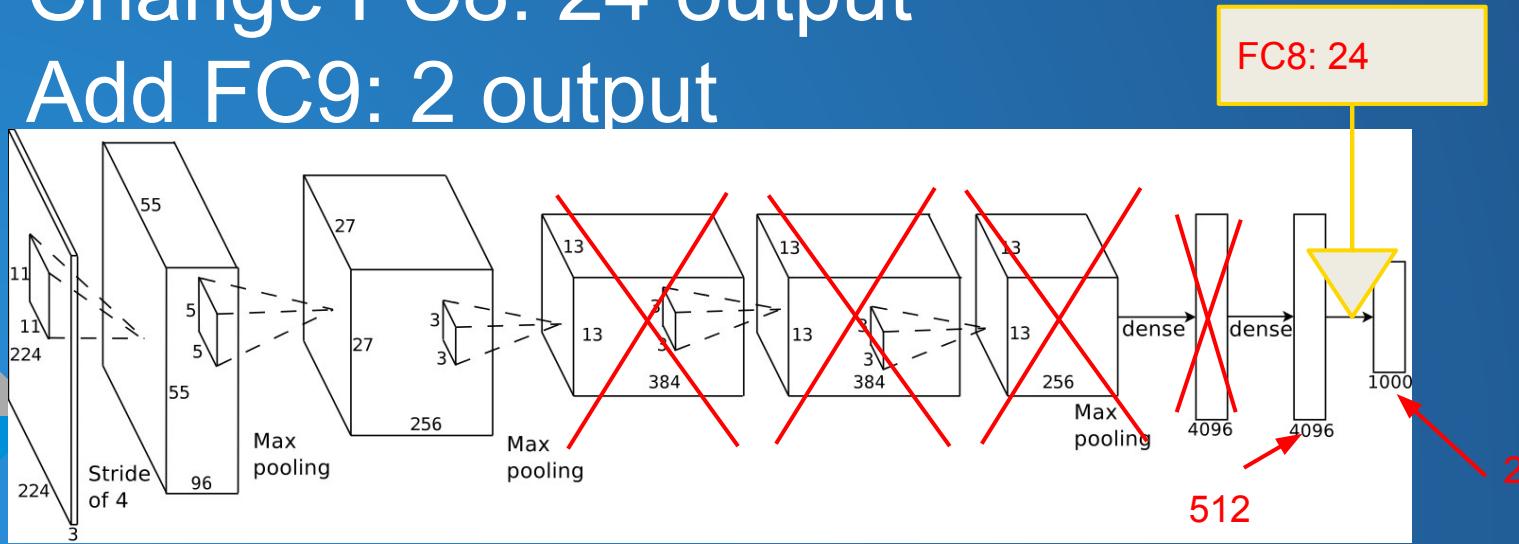
# Lower-Level Model 3

Delete one conv-layer

Reduce FC6 and FC7 4096  $\rightarrow$  512

Change FC8: 24 output

Add FC9: 2 output



# Still Running

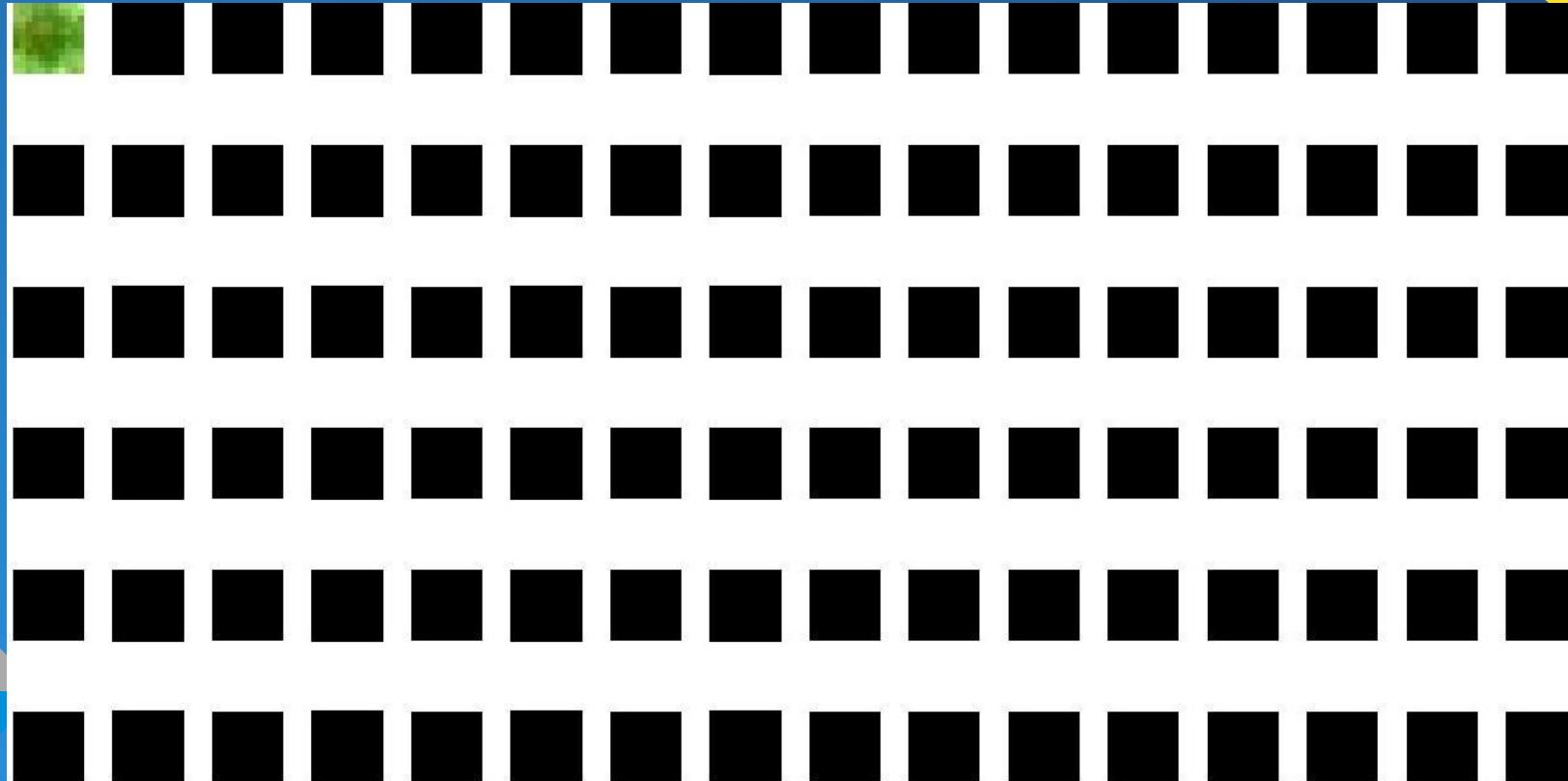
base\_lr: 0.01

stepsize: 100000

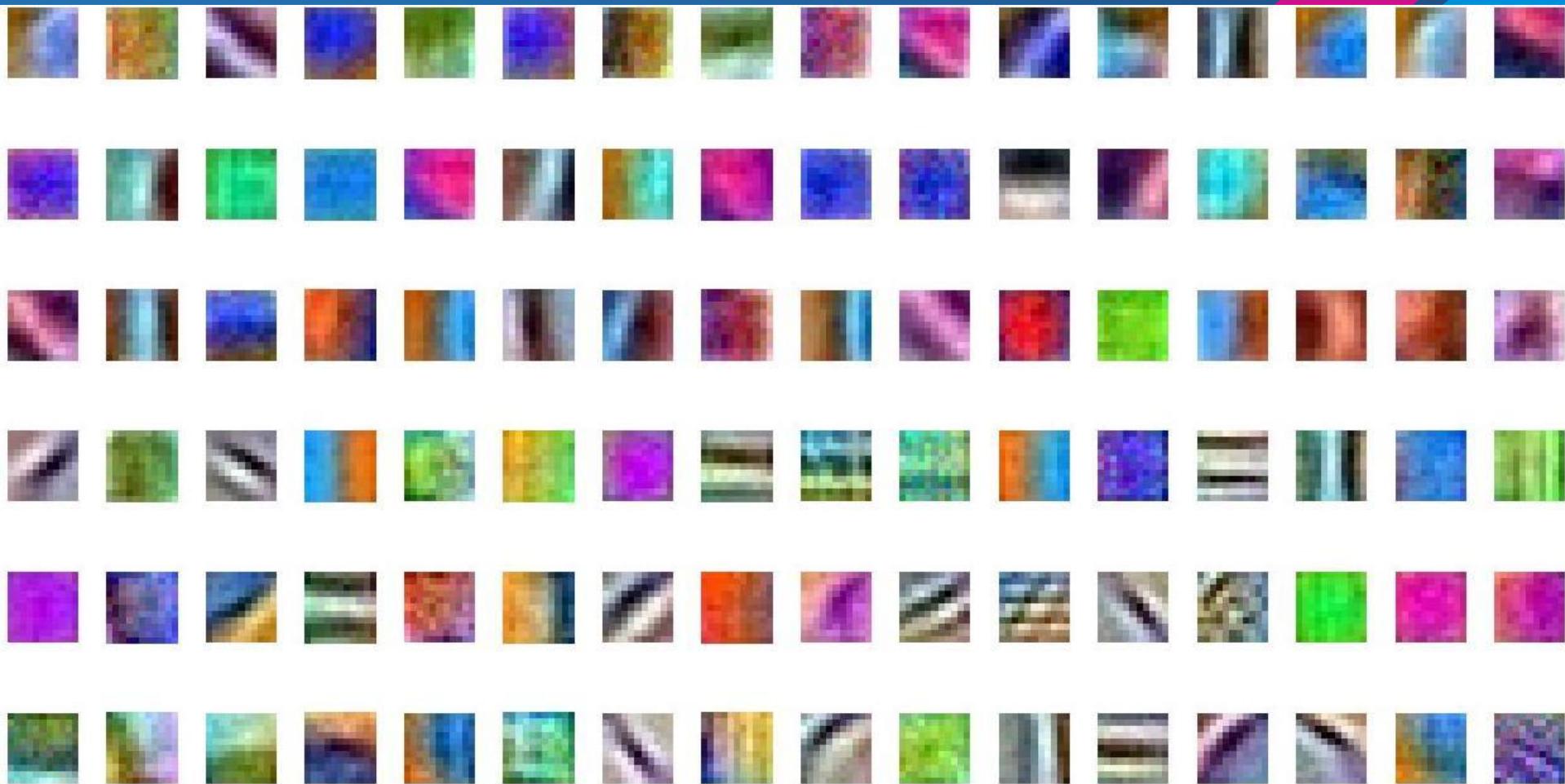
max\_iter: 300000

momentum: 0.9

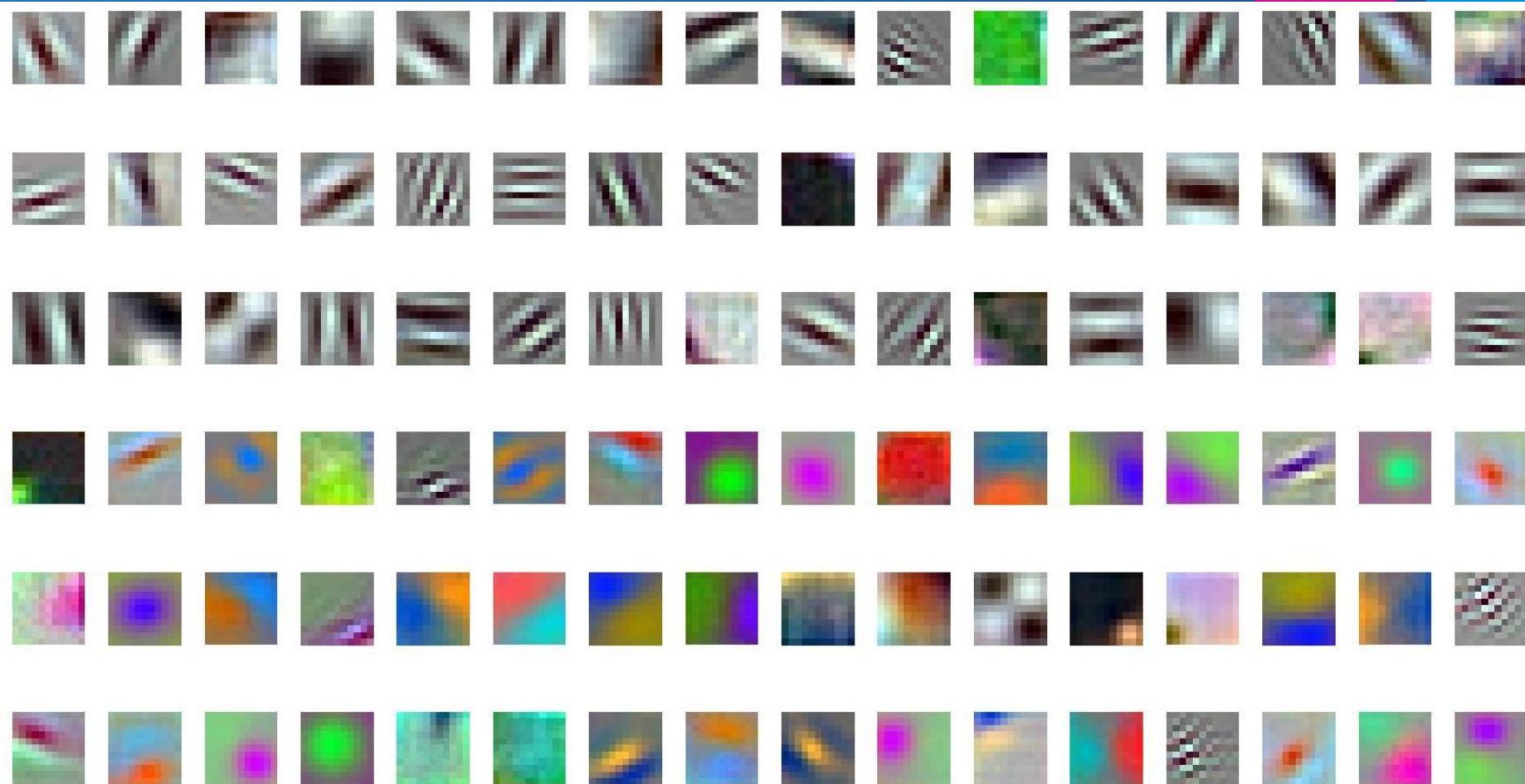
# Without random shuffling 20000



# Still running ( weights on the snapshot of 50000)



# Weights of ImageNet (Pre-trained Model from Caffe)



# 800 Twitter Images

After filtering (animated gif, wrong format)

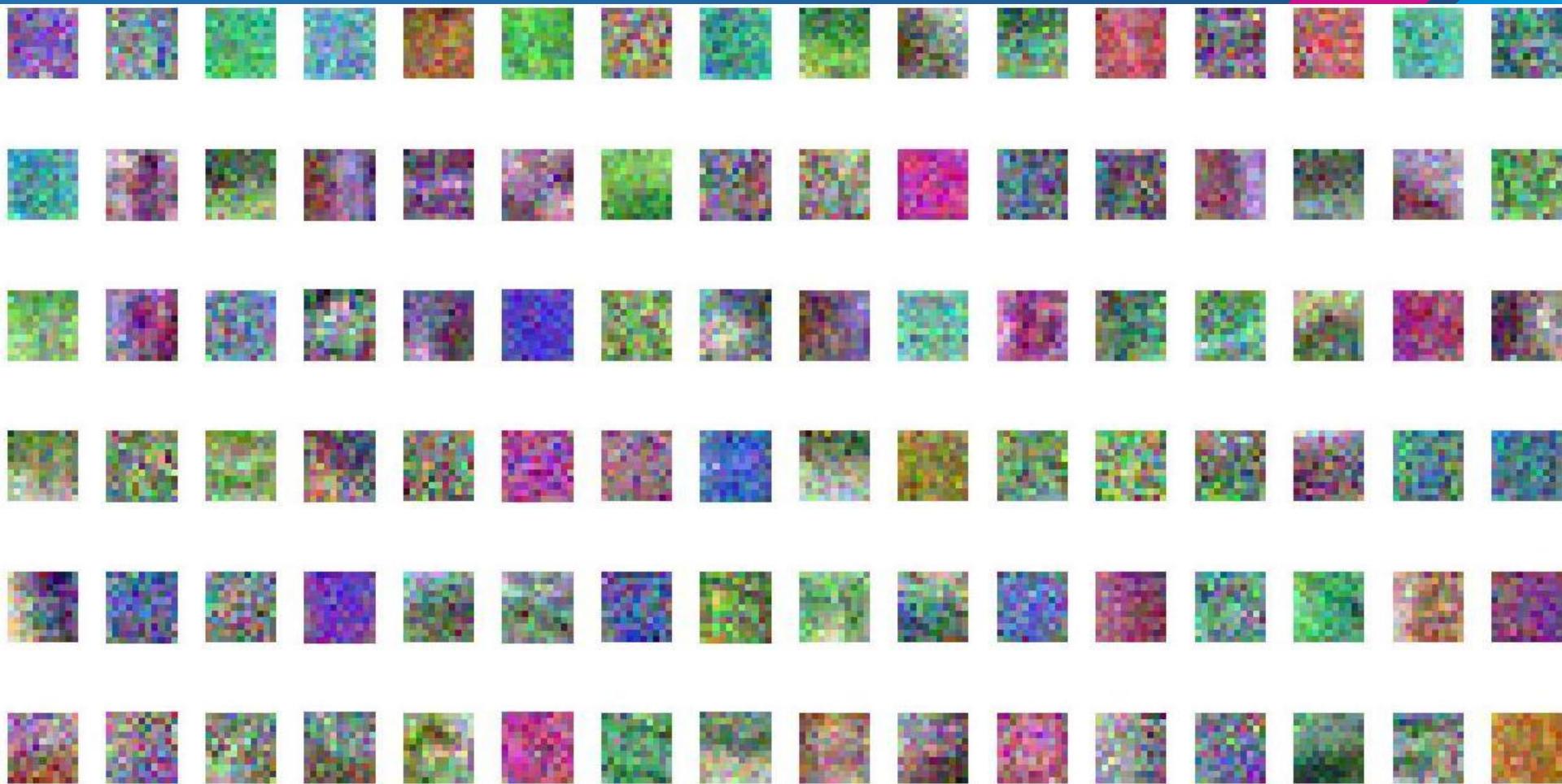
595

Pos:Neg = 462:133 = 0.7765:0.2235

## Same Architecture

100% Accuracy on all images (Overfitting?)

# Weights after 10000 iterations



# Thanks