Supplementary Material
Multimodal Learning with Deep Boltzmann Machines

Nitish Srivastava
Department of Computer Science
University of Toronto
nitish@cs.toronto.edu

Ruslan Salakhutdinov
Department of Statistics and Computer Science
University of Toronto
rsalakhu@cs.toronto.edu

1 Examples of Text Generated by the DBM model

These examples were generated by choosing upto 10 highest probability words under the conditional distribution over the text inputs given the images.

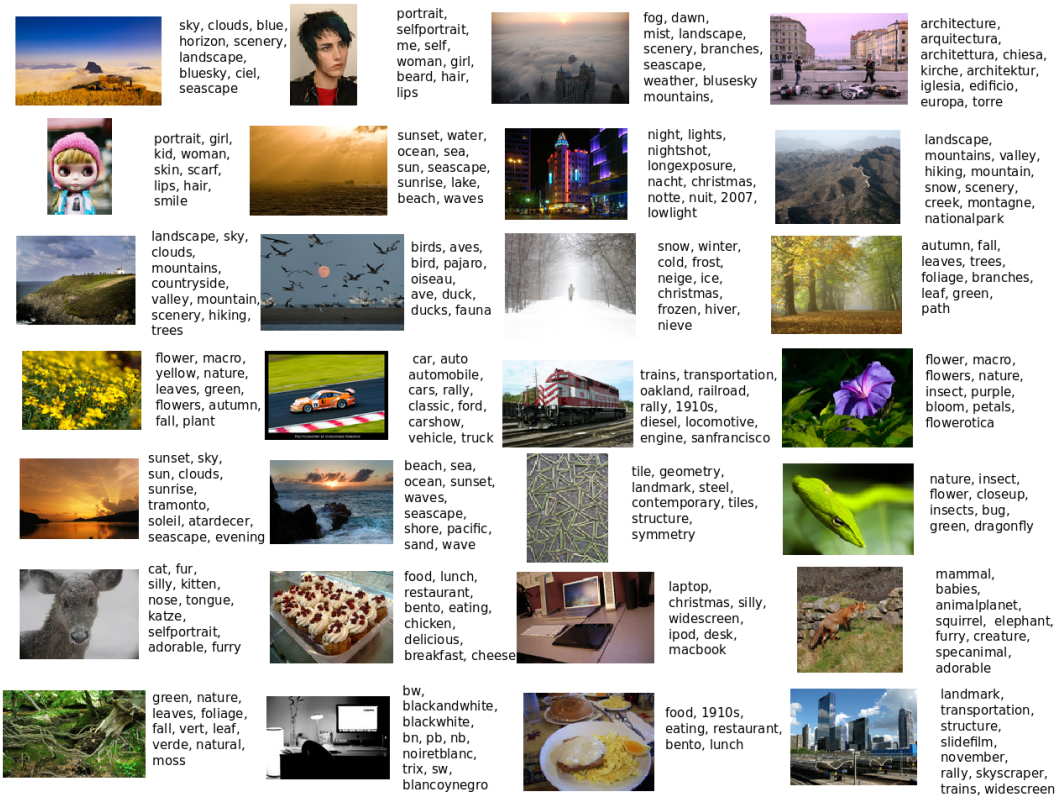

Figure 1: Examples of text generated from images by the DBM model

2 Topic-wise classification results

The classification task considered in this paper involves classifying inputs into topics. There are 38 topics and each input may belong to several topics. In order to make our results comparable

Table 1: Comparison of Average Precision scores of different models on the MIR Flickr Dataset with multimodal inputs.

LABELS	ANIMALS	BABY	BABY*	BIRD	BIRD*	CAR	CAR*	CLOUDS	CLOUDS*	DOG
RANDOM	0.129	0.010	0.005	0.030	0.019	0.047	0.015	0.148	0.054	0.027
LDA	0.537	**0.285**	**0.308**	0.426	0.500	0.297	0.389	0.651	0.528	0.621
SVM	0.531	0.200	0.165	0.443	0.520	0.339	0.434	0.695	0.434	0.607
DBM-LABELLED	0.592	0.101	0.102	0.329	0.343	0.327	0.416	0.751	0.690	0.536
DBM-UNLABELLED	0.667	0.142	0.133	0.473	0.530	0.402	0.539	0.781	0.737	0.692
DBN	**0.688**	0.158	0.150	0.516	0.586	0.417	0.570	0.789	0.748	0.697
AUTOENCODER	0.687	0.162	0.151	0.521	0.594	0.419	0.576	0.789	0.749	0.696
DBM	0.684	0.173	0.167	**0.538**	**0.633**	**0.431**	**0.609**	**0.800**	**0.762**	**0.698**
LABELS	DOG*	FEMALE	FEMALE*	FLOWER	FLOWER*	FOOD	INDOOR	LAKE	MALE	MALE*
RANDOM	0.024	0.247	0.159	0.073	0.043	0.040	0.333	0.032	0.243	0.146
LDA	0.663	0.494	0.454	0.560	0.623	0.439	0.663	0.258	0.434	0.354
SVM	0.641	0.465	0.451	0.480	0.717	0.308	0.683	0.207	0.413	0.335
DBM-LABELLED	0.504	0.582	0.541	0.621	0.711	0.506	0.741	0.245	0.520	0.411
DBM-UNLABELLED	0.695	0.611	0.575	0.674	0.788	0.596	0.769	0.287	0.549	0.451
DBN	0.705	**0.619**	**0.584**	0.683	0.802	0.623	0.776	0.298	0.558	0.465
AUTOENCODER	0.706	0.619	0.583	0.684	0.802	0.630	0.775	0.300	**0.559**	0.467
DBM	**0.715**	0.615	0.578	**0.689**	**0.817**	**0.652**	**0.778**	**0.311**	0.557	**0.472**
LABELS	NIGHT	NIGHT*	PEOPLE	PEOPLE*	PLANT_LIFE	PORTRAIT	PORTRAIT*	RIVER	RIVER*	SEA
RANDOM	0.108	0.027	0.415	0.314	0.351	0.157	0.153	0.036	0.006	0.053
LDA	0.615	0.420	0.731	0.664	0.703	0.543	0.541	0.317	**0.134**	0.477
SVM	0.588	0.450	0.748	0.565	0.691	0.480	0.558	0.158	0.109	0.529
DBM-LABELLED	0.674	0.500	0.819	0.764	0.804	0.668	0.664	0.219	0.042	0.564
DBM-UNLABELLED	0.701	0.565	0.848	0.799	0.823	0.700	0.697	0.283	0.053	0.650
DBN	0.705	0.577	**0.854**	**0.807**	0.829	**0.708**	**0.706**	0.303	0.057	0.659
AUTOENCODER	0.704	0.577	0.853	0.806	0.828	0.707	0.705	0.304	0.057	0.661
DBM	**0.708**	**0.591**	0.846	0.798	**0.831**	0.701	0.699	**0.322**	0.069	**0.677**
LABELS	SEA*	SKY	STRUCTURES	SUNSET	TRANSPORT	TREE	TREE*	WATER	**MEAN**	
RANDOM	0.009	0.316	0.400	0.085	0.116	0.187	0.027	0.133	0.124	
LDA	0.197	0.800	0.709	0.528	0.411	0.515	0.342	0.575	0.492	
SVM	0.201	0.823	0.695	0.613	0.369	0.559	0.321	0.527	0.475	
DBM-LABELLED	0.183	0.862	0.783	0.649	0.456	0.665	0.449	0.650	0.526	
DBM-UNLABELLED	0.253	0.878	0.807	0.668	0.504	0.690	0.524	0.711	0.585	
DBN	0.266	0.883	0.813	0.674	0.521	0.700	0.536	0.724	0.599	
AUTOENCODER	0.271	0.883	0.812	0.675	0.523	0.701	0.539	0.725	0.600	
DBM	**0.298**	**0.888**	**0.816**	**0.683**	**0.535**	**0.710**	**0.555**	**0.733**	**0.609**	

Table 2: Comparison of Average Precision scores of different models with unimodal inputs from the MIR Flickr Dataset.

LABELS	ANIMALS	BABY	BABY*	BIRD	BIRD*	CAR	CAR*	CLOUDS	CLOUDS*	DOG
IMAGE-SVM	0.278	0.084	0.088	0.128	0.129	0.179	0.227	0.651	0.511	0.155
IMAGE-DBN	0.500	0.063	0.057	0.184	0.175	0.261	0.342	0.710	0.640	0.386
IMAGE-DBM	0.513	0.065	0.063	0.203	0.196	0.266	0.349	0.713	0.644	0.401
DBM-ZEROTEXT	0.587	0.100	0.100	0.319	0.333	0.322	0.409	0.749	0.687	0.523
DBM-GENTEXT	**0.599**	**0.105**	**0.101**	**0.343**	**0.362**	**0.333**	**0.428**	**0.753**	**0.694**	**0.556**
LABELS	DOG*	FEMALE	FEMALE*	FLOWER	FLOWER*	FOOD	INDOOR	LAKE	MALE	MALE*
IMAGE-SVM	0.156	0.461	0.389	0.469	0.519	0.293	0.605	0.188	0.407	0.294
IMAGE-DBN	0.350	0.548	0.499	0.539	0.608	0.442	0.702	0.203	0.487	0.373
IMAGE-DBM	0.367	0.550	0.502	0.551	0.624	0.449	0.705	0.204	0.488	0.374
DBM-ZEROTEXT	0.491	0.581	0.539	0.617	0.705	0.502	0.739	0.242	0.518	0.409
DBM-GENTEXT	**0.526**	**0.583**	**0.543**	**0.626**	**0.721**	**0.511**	**0.742**	**0.248**	**0.521**	**0.413**
LABELS	NIGHT	NIGHT*	PEOPLE	PEOPLE*	PLANT_LIFE	PORTRAIT	PORTRAIT*	RIVER	RIVER*	SEA
IMAGE-SVM	0.554	0.390	0.631	0.558	0.687	0.493	0.493	0.179	**0.102**	0.366
IMAGE-DBN	0.629	0.427	0.784	0.721	0.777	0.619	0.613	0.168	0.033	0.490
IMAGE-DBM	0.632	0.432	0.786	0.724	0.777	0.624	0.619	0.170	0.030	0.496
DBM-ZEROTEXT	0.672	0.495	0.817	0.762	0.803	0.666	0.662	0.215	0.042	0.558
DBM-GENTEXT	**0.676**	**0.508**	**0.820**	**0.766**	**0.804**	**0.670**	**0.667**	**0.224**	0.041	**0.573**
LABELS	SEA*	SKY	STRUCTURES	SUNSET	TRANSPORT	TREE	TREE*	WATER	**MEAN**	
IMAGE-SVM	0.126	0.775	0.626	0.588	0.298	0.514	0.205	0.448	0.375	
IMAGE-DBN	0.141	0.835	0.748	0.595	0.406	0.629	0.325	0.569	0.463	
IMAGE-DBM	0.145	0.835	0.750	0.602	0.411	0.632	0.342	0.579	0.469	
DBM-ZEROTEXT	0.180	0.861	0.781	0.647	0.454	0.663	0.441	0.645	0.522	
DBM-GENTEXT	**0.189**	**0.862**	**0.784**	**0.650**	**0.460**	**0.666**	**0.459**	**0.656**	**0.531**	

to previous results, we do one-vs-all classification for each topic separately. For a more detailed analysis of the results, Table 1 reports class-wise average precision scores for the models described in the paper with multimodal inputs.

Table 2 reports class-wise average precision scores for the models with unimodal inputs.