Variational learning across domains with triplet information

Rita Kuznetsova^{1,2}**, Oleg Bakhteev**^{1,2} **and Alexandr Ogaltsov**^{2,3}
^{1}Moscow Institute of Physics and Technology
^{2}Antiplagiat Company
^{3}National Research University Higher School of Economics
{rita.kuznetsova, bakhteev}@phystech.edu, avogaltsov@edu.hse.ru

Abstract

The work investigates deep generative models, which allow us to use training data from one domain to build a model for another domain. We propose the Variational Bi-domain Triplet Autoencoder (VBTA) that learns a joint distribution of objects from different domains. We extend the VBTAs objective function by the relative constraints or triplets that sampled from the shared latent space across domains. In other words, we combine the *deep generative models* with a *metric learning* ideas in order to improve the final objective with the triplets information. The performance of the VBTA model is demonstrated on different tasks: image-to-image translation, bi-directional image generation and cross-lingual document classification.

1 Introduction

Inspired by works Karaletsos et al. [2015], Kingma et al. [2014], Suzuki et al. [2016], Vedantam et al. [2017] we propose Variational Bi-domain Triplet Autoencoder (VBTA) that learns a joint distribution of objects from different domains $\mathbf{X}$ and $\mathbf{Y}$ having a similar structure (e.g. texts, images) with the help of probabilistic triplet modeling. But, unlike these works, we suppose other form of approximate posterior distributions and sampled third triplet object across domains during training process. We suppose that on each training epoch the information from the triplets regularizes our objective. VBTA allows using distributed representations as samples from shared latent space $\mathbf{z}$ that captures characteristics from both domains. We make assumptions about shared-latent space, in which the paired objects (images, sentences) from different domains are close to each other. The main contributions of this paper are the following:

- We introduce the Variational Bi-domain Triplet Autoencoder (VBTA) — new extension of variational autoencoder that trains a joint distribution of objects across domains with learning triplet information. We propose negative sampling method that samples from the shared latent space purely unsupervised during training.
- We demonstrate the performance of the proposed model on different tasks such as bi-directional image generation, image-to-image translation, cross-lingual document classification.

2 Assumptions

Consider dataset $(\mathbf{X}, \mathbf{Y}) = \{\mathbf{x}, \mathbf{y}\}_{n=1}^{N}$ consisting of N *i.i.d.* objects from different domains. We assume that these objects are generated independently by the random process using the same latent variable $\mathbf{z}$. We make an assumption that for each pair $(\mathbf{x}, \mathbf{y})$ there exists a shared latent space variable $\mathbf{z}$, from which we can reconstruct both $\mathbf{x}$ and $\mathbf{y}$. Latent space variable $\mathbf{z}$ is built from the domain space

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

variables $\mathbf{h}_x$, $\mathbf{h}_y$ according to equations: $\mathbf{z} = E(\mathbf{h}_{x_i}) = E\left(E_x(\mathbf{x}_i)\right)$, $\mathbf{z} = E(\mathbf{h}_{y_j}) = E\left(E_y(\mathbf{y}_j)\right)$, where $\mathbf{h}_{x_i}$ and $\mathbf{h}_{y_j}$ are produced from $\mathbf{x}_i$ and $\mathbf{y}_j$ accordingly: $\mathbf{h}_{x_i} = E_x(\mathbf{x}_i)$, $\mathbf{h}_{y_j} = E_y(\mathbf{y}_j)$. We define a shared intermediate variable $\mathbf{h}$, which is used to obtain corresponding domain variables $\hat{\mathbf{x}}_i$, $\hat{\mathbf{y}}_j$ from $\mathbf{y}_j$, $\mathbf{x}_i$ through $\mathbf{z}$: $\mathbf{h} = D(\mathbf{z}) = D\left(E(E_x(\mathbf{x}_i))\right) = D\left(E(E_y(\mathbf{y}_j))\right)$.

$\hat{\mathbf{y}}_j = D_y(\mathbf{z}) = D_y\left(D(E(E_x(\mathbf{x}_i)))\right) = f(\mathbf{x}_i) \approx \mathbf{y}_j, \hat{\mathbf{x}}_i = D_x(\mathbf{z}) = D_x\left(D(E(E_y(\mathbf{x}_i)))\right) = g(\mathbf{y}_j) \approx \mathbf{x}_i$. The necessary condition for f and g to exist is the cycle-consistency constraint. That is, the proposed assumptions requires the cycle-consistency assumption. The following diagram on Figure 1 presents VBTA generative process. Objects $\mathbf{z}_i$, $\mathbf{z}_i$ and $\mathbf{z}_k$ form triplet.

$$\begin{tikzcd}
 & & \mathbf{x}_i \arrow[r] \arrow[dl, "E_x"'] \arrow[dd, "E\circ E_x"] & \hat{\mathbf{x}}_i \\
\mathbf{z}_i & \mathbf{h}_{x_i} \arrow[dr, "E"] & & \\
\mathbf{z}_j & & \mathbf{z} \arrow[ull] \arrow[ll] \arrow[dll] \arrow[r, "D"'] \arrow[uur, "D_x\circ D"] \arrow[ddr, "D_y\circ D"] & \mathbf{h} \arrow[uu, "D_x"'] \arrow[dd, "D_y"] \\
\mathbf{z}_k & \mathbf{h}_{y_j} \arrow[ur, "E"] & & \\
 & & \mathbf{y}_j \arrow[r] \arrow[ul, "E_y"] \arrow[uu, "E\circ E_y"'] & \hat{\mathbf{y}}_j
\end{tikzcd}$$

Figure 1: VBTA generative process

3 Variational Bi-domain Triplet Autoencoder

The marginal likelihood defined by this model is:

$$p(\mathbf{x}, \mathbf{y}, \mathbf{t}) = \int_{\mathbf{z}} p_{\theta_{\mathbf{x}}}(\mathbf{x}|\mathbf{z}) p_{\theta_{\mathbf{y}}}(\mathbf{y}|\mathbf{z}) p(\mathbf{t}_{i,j,k}|\mathbf{z}_i, \mathbf{z}_j, \mathbf{z}_k) p(\mathbf{z}) d\mathbf{z} \tag{1}$$

We can assume the following generative process:

- generate $\mathbf{z}$ from prior distribution $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$,
- value $\mathbf{x}$ is generated from some conditional distribution $p_{\theta_{\mathbf{x}}}(\mathbf{x}|\mathbf{z})$,
- value $\mathbf{y}$ is generated from some conditional distribution $p_{\theta_{\mathbf{y}}}(\mathbf{y}|\mathbf{z})$.

The lower bound of the log-likelihood:

$$\begin{aligned}
\mathcal{L}_{VBTA} &= \mathbb{E}_{q_{\phi_x}(\mathbf{z}_x|\mathbf{x})} \log \frac{p_{\theta_x}(\mathbf{x}, \mathbf{y}, \mathbf{t}, \mathbf{z}_x)}{q_{\phi_x}(\mathbf{z}_x|\mathbf{x})} + \mathbb{E}_{q_{\phi_y}(\mathbf{z}_y|\mathbf{y})} \log \frac{p_{\theta_y}(\mathbf{x}, \mathbf{y}, \mathbf{t}, \mathbf{z}_y)}{q_{\phi_y}(\mathbf{z}_y|\mathbf{y})} = \\
&= \underbrace{-\Big[KL\big(q_{\phi_{\mathbf{x}}(\mathbf{z}_x|\mathbf{x})}(\mathbf{z}_x|\mathbf{x}) \parallel p_{\theta_{\mathbf{x}}}(\mathbf{z}_x)\big) + KL\big(q_{\phi_{\mathbf{y}}(\mathbf{z}_y|\mathbf{y})}(\mathbf{z}_y|\mathbf{y}) \parallel p_{\theta_{\mathbf{y}}}(\mathbf{z}_y)\big)\Big]}_{\text{Penalty}} + \\
&\quad + \underbrace{\Big[\mathbb{E}_{q_{\phi_{\mathbf{x}}}(\mathbf{z}_x|\mathbf{x})}\big[\log p_{\theta_{\mathbf{x}}}(\mathbf{x}|\mathbf{z}_x)\big] + \mathbb{E}_{q_{\phi_{\mathbf{y}}}(\mathbf{z}_y|\mathbf{y})}\big[\log p_{\theta_{\mathbf{y}}}(\mathbf{y}|\mathbf{z}_y)\big]\Big]}_{\text{Reconstruction}} + \\
&\quad + \underbrace{\Big[\mathbb{E}_{q_{\phi_{\mathbf{x}}}(\mathbf{z}_x|\mathbf{x})}\big[\log p_{\theta_{\mathbf{x}}}(\mathbf{y}|\mathbf{z}_x)\big] + \mathbb{E}_{q_{\phi_{\mathbf{y}}}(\mathbf{z}_y|\mathbf{y})}\big[\log p_{\theta_{\mathbf{y}}}(\mathbf{x}|\mathbf{z}_y)\big]\Big]}_{\text{Cycle-consistency}} + \\
&\quad + \underbrace{\mathbb{E}_{q_{\phi_{\mathbf{x}}}(\mathbf{z}_x|\mathbf{x})}\big[\log p(\mathbf{t}|\mathbf{z}_x)\big] + \mathbb{E}_{q_{\phi_{\mathbf{y}}}(\mathbf{z}_y|\mathbf{x})}\big[\log p(\mathbf{t}|\mathbf{z}_y)\big]}_{\text{Triplet likelihood}}
\end{aligned} \tag{2}$$

Both $q_{\phi_{\mathbf{x}}(\mathbf{z}_x|\mathbf{x})}(\mathbf{z}_x|\mathbf{x})$ and $q_{\phi_{\mathbf{y}}(\mathbf{z}_y|\mathbf{y})}(\mathbf{z}_y|\mathbf{y})$ are encoders, $p_{\theta_{\mathbf{x}}}(\mathbf{x}|\mathbf{z}_x)$ and $p_{\theta_{\mathbf{y}}}(\mathbf{y}|\mathbf{z}_y)$ are decoders, modeled by the deep neural networks. Similar to Liu et al. [2017] our decoders and encoders use the common functions E and D, see (2). We apply the Stochastic Gradient Variational Bayes (SGVB) and optimize the variational parameters $\theta_{\mathbf{x}}$, $\theta_{\mathbf{y}}$, $\phi_{\mathbf{x}}$ and $\phi_{\mathbf{y}}$.

4 Learning Triplets

Based on the metric learning approach and similar to Karaletsos et al. [2015] we extend our model by relative constraints or triplets: $\mathcal{T} = \{(\mathbf{z}_i, \mathbf{z}_j, \mathbf{z}_k) : d(\mathbf{z}_i, \mathbf{z}_j) < d(\mathbf{z}_i, \mathbf{z}_k)\}$, but in our case we sampled triplets across domains $\mathbf{X}$ and $\mathbf{Y}$. We define the conditional triplet likelihood in the following form:

$$p(t_{i,j,k} = True|i, j, k) = \int_{\mathbf{z}} p(\mathbf{t}_{i,j,k}|\mathbf{z}_i, \mathbf{z}_j, \mathbf{z}_k)p(\mathbf{z}_i)p(\mathbf{z}_j)p(\mathbf{z}_k)d\mathbf{z}_i d\mathbf{z}_j d\mathbf{z}_k, \tag{3}$$

that was modelled by Bernoulli distribution over the states *True* and *False* parametrized with the use of softmax-function

$$p(t_{i,j,k}|i, j, k) = \frac{e^{-d(\mathbf{z}_i, \mathbf{z}_j)}}{e^{-d(\mathbf{z}_i, \mathbf{z}_j)} + e^{-d(\mathbf{z}_i, \mathbf{z}_k)}} \tag{4}$$

Triplets — three objects from shared latent space $\mathbf{z}$. $\mathbf{z}_i$, $\mathbf{z}_j$ — shared latent representation of objects from $\mathbf{X}$ and $\mathbf{Y}$ domains. The third object $\mathbf{z}_k$ is sampled from domain $\mathbf{y}$ with the minimal distance function to the corresponding objects from domain $\mathbf{x}$ (and vice versa):

$$\mathbf{z}_k = \underset{\mathbf{z}_{i'} \in \mathcal{S}_b \setminus (\mathbf{z}_i, \mathbf{z}_j)}{\arg\min} d(\mathbf{z}_i, \mathbf{z}_{i'}), \tag{5}$$

where $\mathcal{S}_b \in \mathcal{S}$ — current mini-batch, $\mathbf{z}_i$ and $\mathbf{z}_j$ — the paired objects from different domains. As d we use approximate form of JS-divergence, like Karaletsos et al. [2015]. In other words, we want to choose an example $\mathbf{z}_k$ that is similar to $\mathbf{z}_i$ according to the current model parameters.

5 Experiment and Results

We presented the results on an image-to-image translation task: MNIST LeCun et al. [1998] and CelebA Liu et al. [2015]. We presented results on cross-lingual text classification task on RCV1/RCV2 corpora Lewis et al. [2004]. See architecture details in (6).

5.1 Image-to-image translation for MNIST dataset

We evaluated our approach on MNIST-transpose, where the two image domains $\mathbf{x}$ and $\mathbf{y}$ are the MNIST images and their corresponding transposed ones. Similar to Gan et al. [2017] we used the classifier that trained on MNIST images as a ground-truth evaluator. For all the transposed images we encoded them via the model encoder $E \circ E_{\mathbf{y}}$ and decoded via decoder $D_{\mathbf{x}} \circ D$. Then we sent classified it. The results of the classification are shown in Table 1, where n is the number of objects used for triplets sampling and cycle-consistency.

Table 1: Classification accuracy (%) on the MNIST-transpose dataset. The DiscoGAN, Triple GAN and Δ-GAN results are taken from Gan et al. [2017]

Model	$n = 0$	$n = 10$	$n = 100$	$n = 1000$	All
DiscoGAN	-	-	-	-	15.00 ± 0.20
Triple GAN	-	-	63.79 ± 0.85	84.93 ± 1.63	86.70 ± 1.52
Δ-GAN	-	-	83.20 ± 1.88	88.98 ± 1.50	93.34 ± 1.46
VBTA	18.89 ± 3.59	86.57 ± 6.338	$\mathbf{90.44 \pm 0.003}$	$\mathbf{90 \pm 0.0026}$	$\mathbf{95 \pm 0.0006}$

We evaluated the marginal log-likelihood of our model on binarized versions of MNIST and MNIST-transpose. The results are listed in Table 2.

5.2 Qualitative results for CelebA dataset

In this section we considered this dataset as a union of two domains: faces of men $\mathbf{X}$ and faces of women $\mathbf{Y}$. Figure 2 shows the face images from datasets and their translation into different domains.

See example of bi-directional image generation in (6).

Table 2: Marginal log-likelihood for MNIST as log$p(\mathbf{x})$ and MNIST-transpose datasets as log$p(\mathbf{y})$.

Model	$< logp(\mathbf{x})$	$< logp(\mathbf{y})$
VAE Kingma and Welling [2013]	-81.13	-81.01
JVMAE Suzuki et al. [2016]	-85.35	-85.44
VBTA	$\mathbf{-80.92}$	$\mathbf{-80.91}$

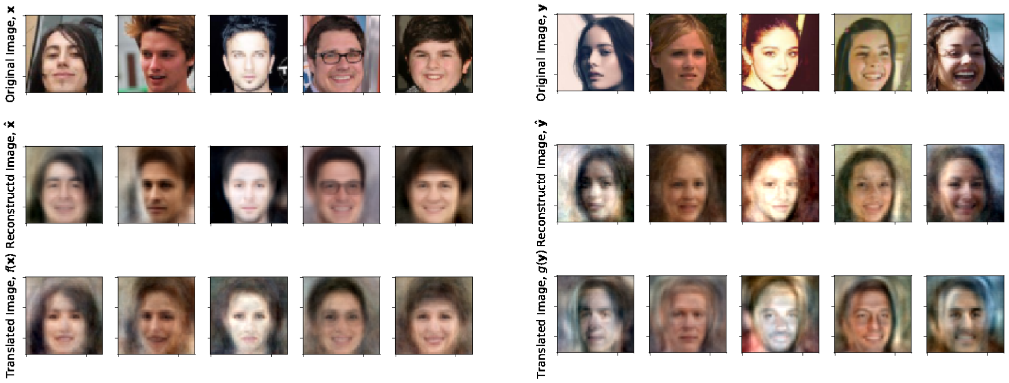

Figure 2: Results of image-to-image translation for CelebA dataset. The first row corresponds to the original images that were considered as similar because of high amount of matching attrbutes. The second row shows the reconstruction of the images. The third row illustrates the image translation from domain $\mathbf{X}$ into domain $\mathbf{Y}$ and from $\mathbf{Y}$ into $\mathbf{X}$.

5.3 Cross Lingual Document Classification

Given a classifier trained on documents in language *A* ($\mathbf{X}$ domain), one should use that classifier to predict labels of documents in language *B* ($\mathbf{Y}$ domain). To handle this task we need to construct meaningful bilingual text representations. For train and test we used RCV1/RCV2 corpora, where documents are assigned to one of four predefined topics. In contrast to previous work, we *do not* use parallel data at all. We artificially paired documents according to their topics. For classification experiment, 10000 documents in English was used to train classifier and test it on 5000 documents in German and vice versa. Classification results are in Table 3. See another details in (6).

Table 3: Text classification accuracy

Model	$en \to de$	$de \to en$
Majority Baseline	46.8	46.8
MT Baseline	68.1	67.4
Klementiev et al. [2012]	77.6	71.1
Gouws et al. [2015]	86.5	75.0
Chandar et al. [2014]	91.8	74.2
Wei and Deng [2017]	92.7	80.4
Su et al. [2018]	91.3	77.8
VBTA	**94.3**	**82.8**

6 Conclusion

In this paper we proposed the Variational Bi-domain Triplet Autoencoder (VBTA) that learns a joint distribution of objects from different domains with the help of the learning triplets that sampled from the shared latent space across domains. We demonstrated the performance of the VBTA model on different tasks: image-to-image translation, bi-directional image generation and cross-lingual document classification.

Acknowledgments. This work was supported by RFBR project No.18-07-01441

References

A. P. S. Chandar, S. Lauly, H. Larochelle, M. M. Khapra, B. Ravindran, V. Raykar, and A. Saha. An autoencoder approach to learning bilingual word representations. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, pages 1853–1861, Cambridge, MA, USA, 2014. MIT Press.

Z. Gan, L. Chen, W. Wang, Y. Pu, Y. Zhang, H. Liu, C. Li, and L. Carin. Triangle generative adversarial networks. *CoRR*, abs/1709.06548, 2017. URL `http://arxiv.org/abs/1709.06548`.

S. Gouws, Y. Bengio, and G. Corrado. Bilbowa: Fast bilingual distributed representations without word alignments. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 748–756. JMLR.org, 2015.

T. Karaletsos, S. Belongie, and G. Rätsch. Bayesian representation learning with oracle constraints. *arXiv preprint arXiv:1506.05011*, 2015.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

D. P. Kingma, S. Mohamed, D. Jimenez Rezende, and M. Welling. Semi-supervised learning with deep generative models. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 3581–3589. 2014.

A. Klementiev, I. Titov, and B. Bhattarai. Inducing crosslingual distributed representations of words. In *COLING*, 2012.

P. Koehn. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Conference Proceedings: the tenth Machine Translation Summit*, pages 79–86, Phuket, Thailand, 2005. AAMT, AAMT.

P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, ACL '07, pages 177–180, Stroudsburg, PA, USA, 2007. Association for Computational Linguistics.

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

D. D. Lewis, Y. Yang, T. G. Rose, and F. Li. Rcv1: A new benchmark collection for text categorization research. *J. Mach. Learn. Res.*, 5:361–397, Dec. 2004.

M. Liu, T. Breuel, and J. Kautz. Unsupervised image-to-image translation networks. *CoRR*, abs/1703.00848, 2017. URL `http://arxiv.org/abs/1703.00848`.

Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3730–3738, 2015.

J. Su, S. Wu, B. Zhang, C. Wu, Y. Qin, and D. Xiong. A neural generative autoencoder for bilingual word embeddings. *Inf. Sci.*, 424(C):287–300, Jan. 2018. ISSN 0020-0255.

M. Suzuki, K. Nakayama, and Y. Matsuo. Joint multimodal learning with deep generative models. *arXiv preprint arXiv:1611.01891*, 2016.

R. Vedantam, I. Fischer, J. Huang, and K. Murphy. Generative models of visually grounded imagination. *arXiv preprint arXiv:1705.10762*, 2017.

L. Wei and Z.-H. Deng. A variational autoencoding approach for inducing cross-lingual word embeddings. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, IJCAI'17, pages 4165–4171, 2017. ISBN 978-0-9992411-0-3.

Appendix A.

Image-to-image translation experiment details

Datasets

We used MNIST dataset for toy problem of image-to-image translation. Similar to Gan et al. [2017] we considered a transposition of this dataset as a second domain $\mathbf{y}$. We used 50,000 as training set and the remaining 10,000 as a test set.

CelebA consists of 202,599 face images with 40 binary attributes. In this work we considered this dataset as a union of two domains: faces of men $\mathbf{x}$ and faces of women $\mathbf{y}$. Similar to Suzuki et al. [2016] we cropped and normalized the images and resized them to 64x64. Since we did not have any paired men and women in CelebA dataset, we considered that the object $\mathbf{y}$ (women) is similar to object $\mathbf{x}$ (men) if they had the largest matching of their attributes.

Model Architecture

For the MNIST dataset we used one-layer network of 512 hidden units with ReLU for decoder D and encoders E_x, E_y. For the modeling shared encoder E and decoder D_x, D_y we used the linear mappings. The shared latent space dimension was set to 64.

For the classification evaluation we set $p_{\theta_{\mathbf{x}}}(\mathbf{x}|\mathbf{z})$ and $p_{\theta_{\mathbf{y}}}(\mathbf{y}|\mathbf{z})$ to be Gaussian distribution. For the comparison to JMVAE Suzuki et al. [2016] model we set $p_{\theta_{\mathbf{x}}}(\mathbf{x}|\mathbf{z})$ and $p_{\theta_{\mathbf{y}}}(\mathbf{y}|\mathbf{z})$ to be Bernoulli. We set model of JMVAE to the same configuration.

For CelebA we used encoders E_x, E_y with two convolution layers and a flattened layer with ReLU. For the shared encoder E and decoder D_x, D_y we used linear mapping into 64 hidden units. For the decoder D we used a network with one dense layer with 8192 units and a deconvolution layer. We considered $p_{\theta_{\mathbf{x}}}(\mathbf{x}|\mathbf{z})$ and $p_{\theta_{\mathbf{y}}}(\mathbf{y}|\mathbf{z})$ as a Gaussian distribution.

We used the Adam Kingma and Ba [2014] optimization algorithm with a learning rate of 10^{-3} for the MNIST dataset and 10^{-4} for CelebA dataset. All the models were trained for 100 epochs with batch size set of 50.

Example of bi-directional image generation

Figure 3 shows faces generated from Gaussian distribution. We found that our algorithm works rather well and can reproduce similar faces for both domains from one sample in latent space.

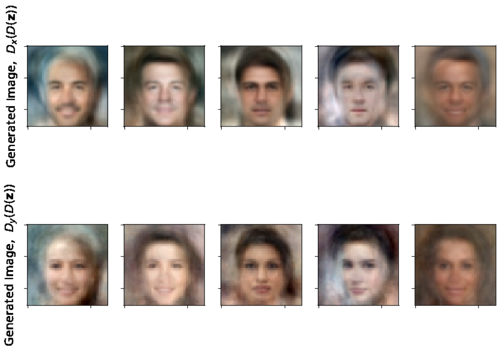
Generated Image, $D_x(D(\mathbf{z}))$
Generated Image, $D_y(D(\mathbf{z}))$

Figure 3: Results of image generation from the common shared space. Each column corresponds to the faces generated from one sample of $\mathbf{z}$. The latent variable $\mathbf{z}$ was sampled from Gaussian distribution: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Cross Lingual Document Classification experiment details

We use experimental setup similar to introduced in Klementiev et al. [2012]. Previous work Chandar et al. [2014], Wei and Deng [2017] and Gouws et al. [2015] used Europarl v7 parallel corpus Koehn [2005] to pretrain embeddings and then utilize it to classify subset of RCV1/RCV2 corpora Lewis et al. [2004]. In this corpora documents are assigned to one of four predefined topics: CCAT (Corporate/Industrial), ECAT (Economics), GCAT (Government/Social), MCAT (Markets). In contrast to previous work, we *do not* use parallel data at all.

We select 15000 documents from both English and German for classification experiments. Algorithm was trained for approximately 300K iterations with batch size equals to 50. We use Moses Koehn et al. [2007] preprocessing tools to lowercase and tokenize texts. Bag-of-words was used as an initial document representation. We keep 30000 top-frequency words for each language as a vocabulary.
We train logistic regression using low-dimensional representation obtained by our algorithm as features.