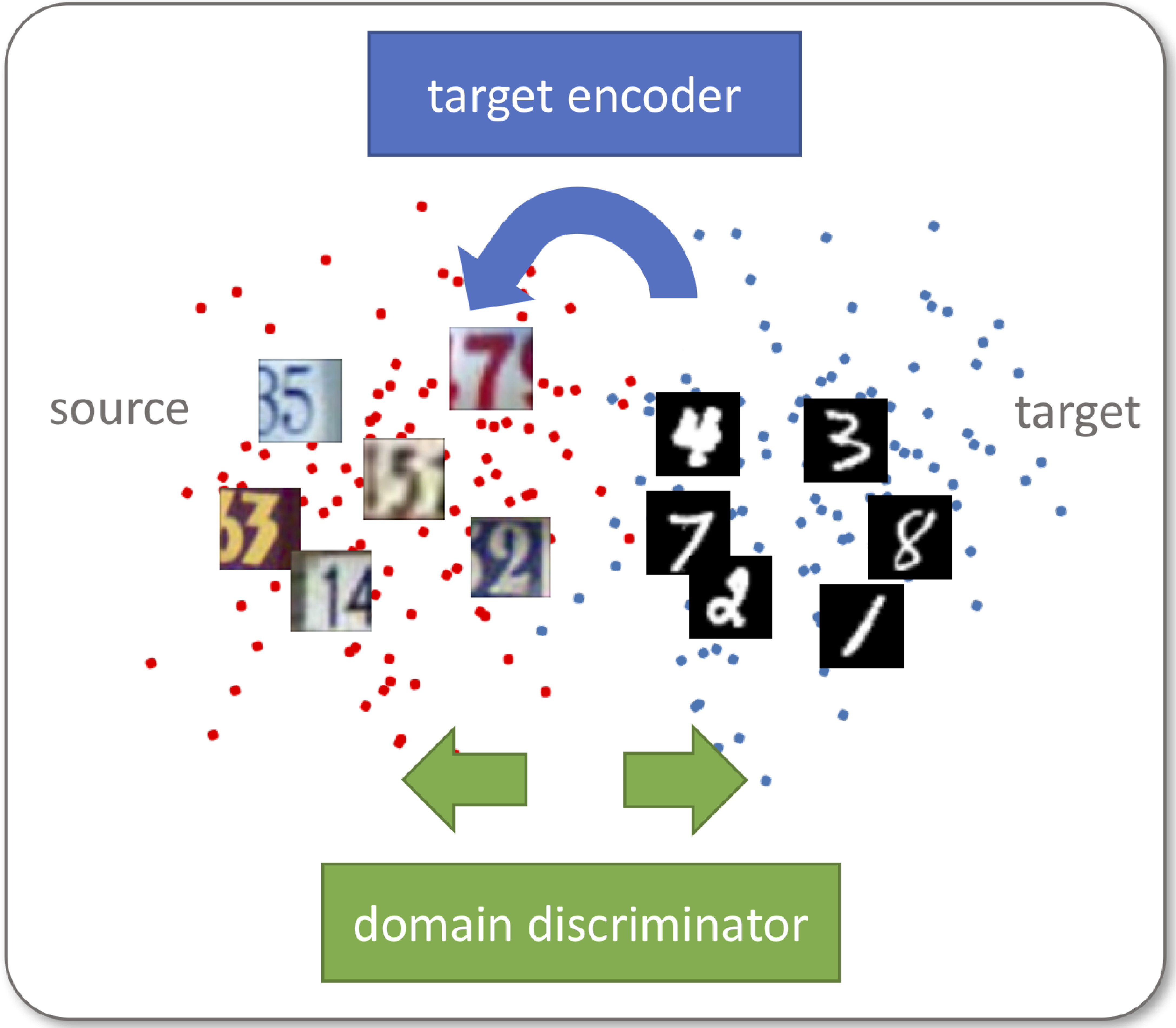
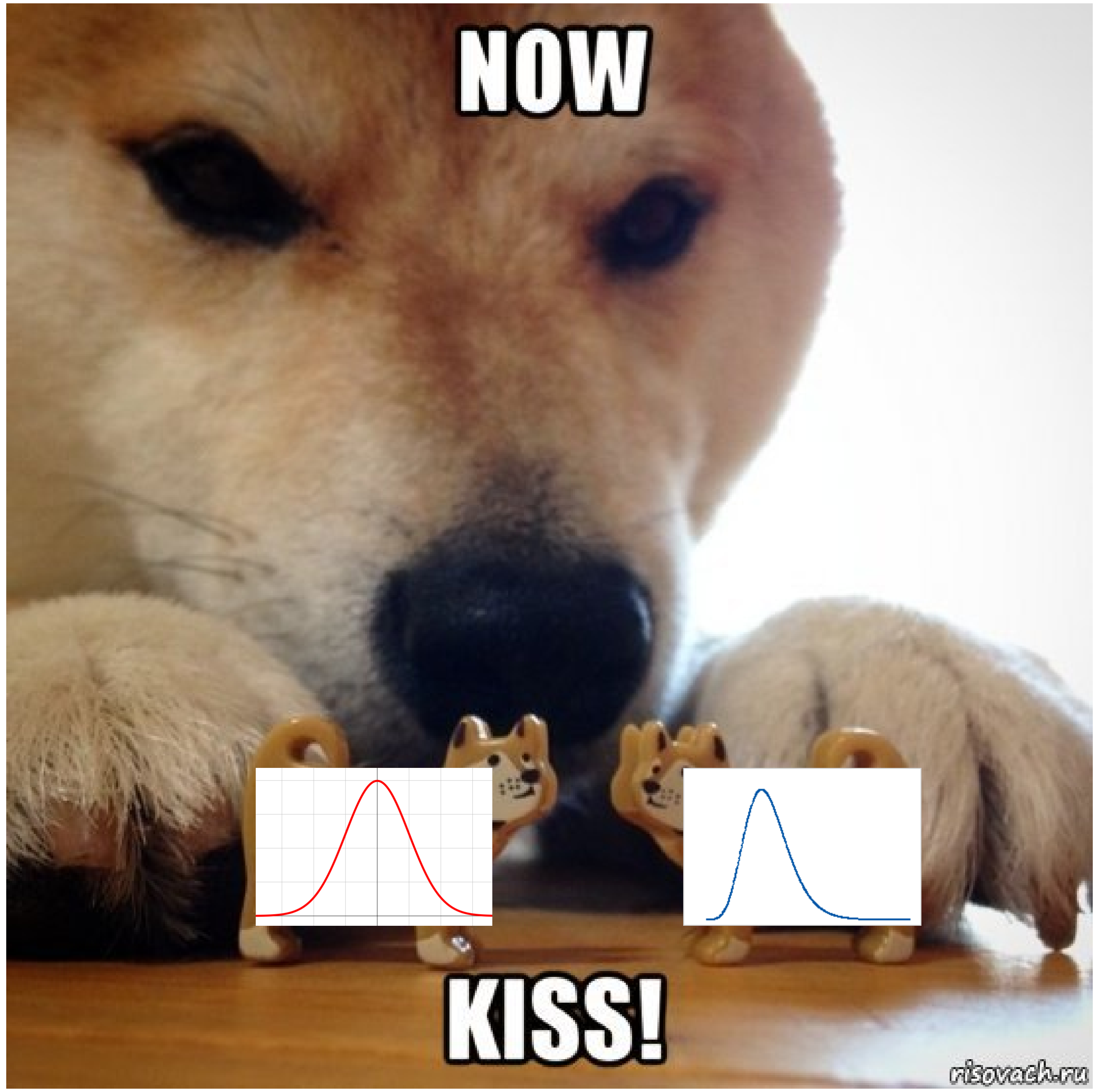
Unsupervised Adversarial Domain Adaptation

What is domain adaptation?



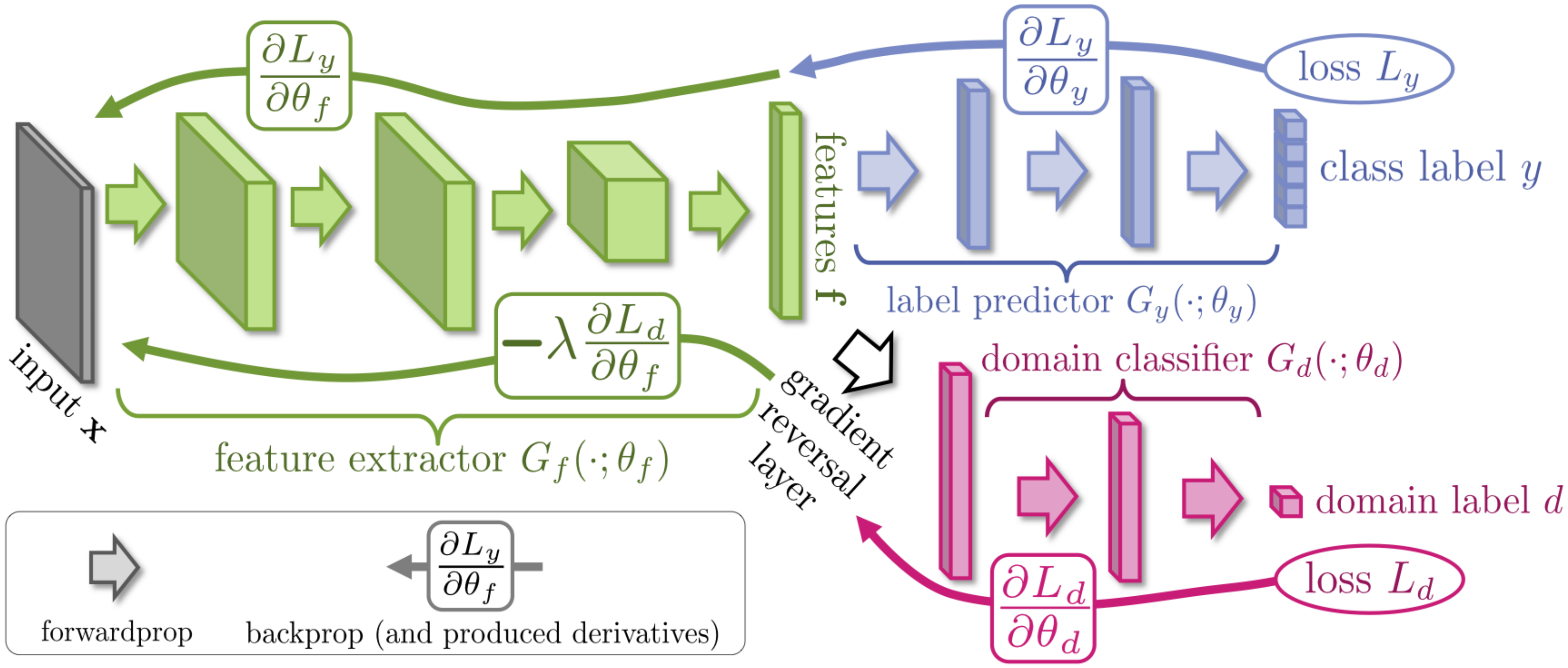
target encoder
source
target
domain discriminator


NOW
KISS!
risovach.ru

DIFFERENT DEEP APPROACHES TO ONE-STEP DA

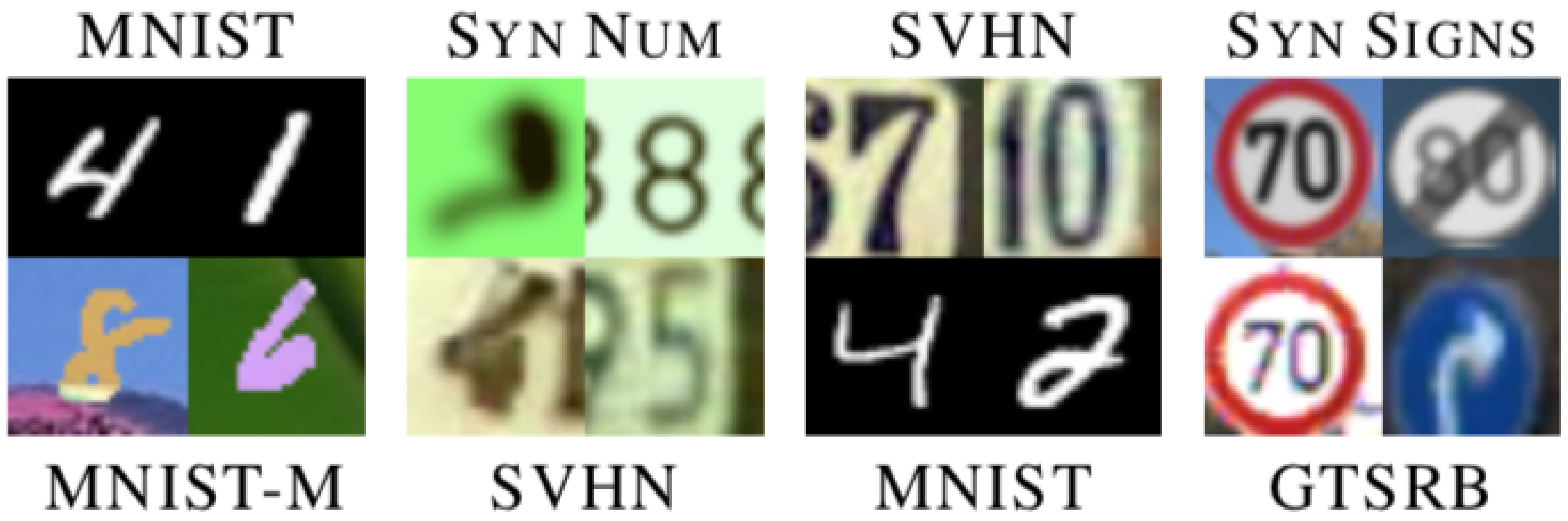
One-step DA Approaches	Brief Description	Subsettings
Discrepancy-based	fine-tuning the deep network with labeled or unlabeled target data to diminish the domain shift	class criterion [118], [86], [79], [98] [53], [45], [75], [139], [130], [29], [118], [28]
		statistic criterion [74], [130], [73] [75], [120], [32], [109], [87], [144]
		architecture criterion [69], [54], [68], [95], [128], [89]
		geometric criterion [16]
Adversarial-based	using domain discriminators to encourage domain confusion through an adversarial objective	generative models [70], [4], [57]
		non-generative models [119], [118], [26], [25], [117] [85]
Reconstruction-based	using the data reconstruction as an auxiliary task to ensure feature invariance	encoder-decoder reconstruction [5], [33], [31], [144]
		adversarial reconstruction [131], [143], [59]

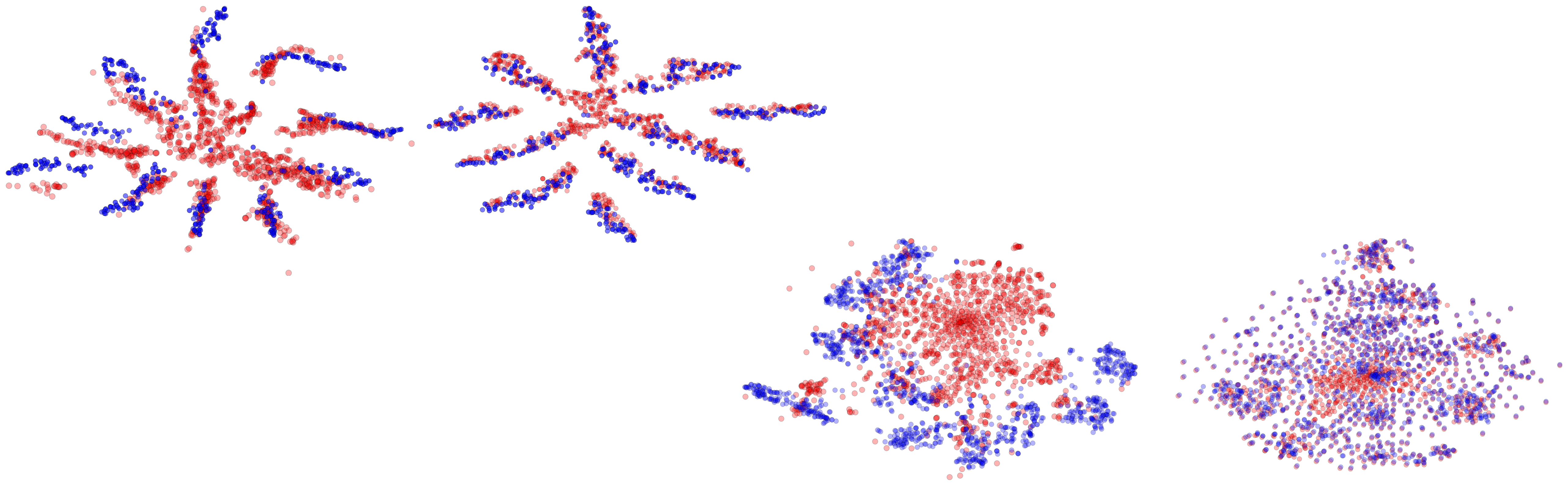
Unsupervised Adversarial Domain Apadatation by Backpropogation



$$
\begin{aligned}
\theta_f &\longleftarrow \theta_f - \mu\left(\frac{\partial L_y^i}{\partial \theta_f} - \lambda \frac{\partial L_d^i}{\partial \theta_f}\right) \\
\theta_y &\longleftarrow \theta_y - \mu \frac{\partial L_y^i}{\partial \theta_y} \\
\theta_d &\longleftarrow \theta_d - \mu \frac{\partial L_d^i}{\partial \theta_d}
\end{aligned}
$$

Nice pictures

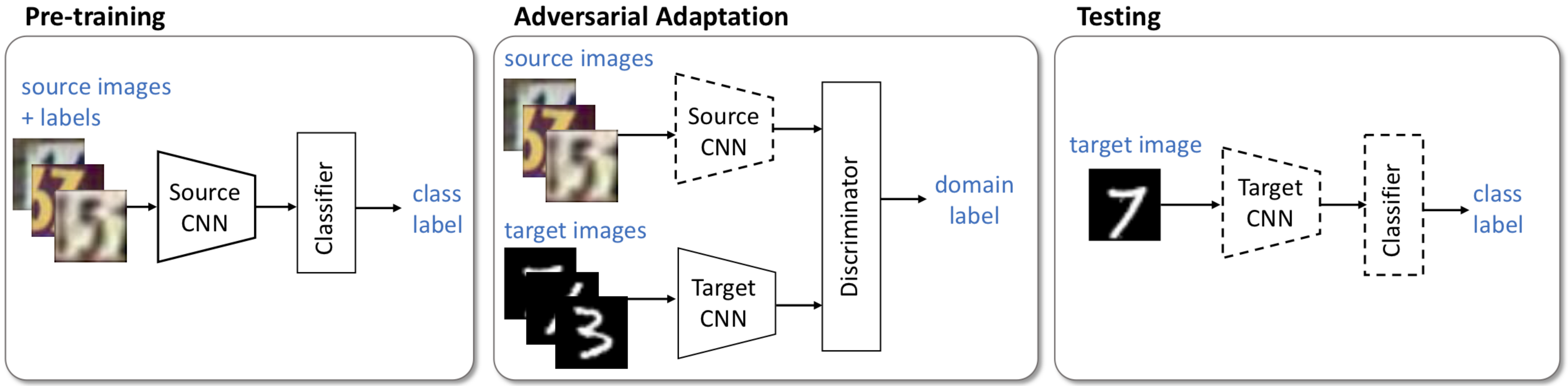




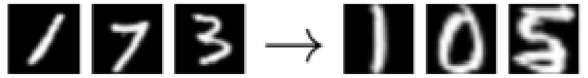
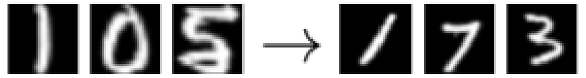
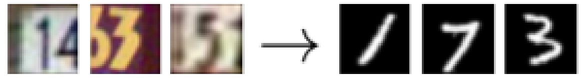
Experiments results

Method	Source	MNIST	Syn Numbers	SVHN	Syn Signs
	Target	MNIST-M	SVHN	MNIST	GTSRB
Source only		.5225	.8674	.5490	.7900
SA (Fernando et al., 2013)		.5690 (4.1%)	.8644 (−5.5%)	.5932 (9.9%)	.8165 (12.7%)
Proposed approach		**.7666** (52.9%)	**.9109** (79.7%)	**.7385** (42.6%)	**.8865** (46.4%)
Train on target		.9596	.9220	.9942	.9980

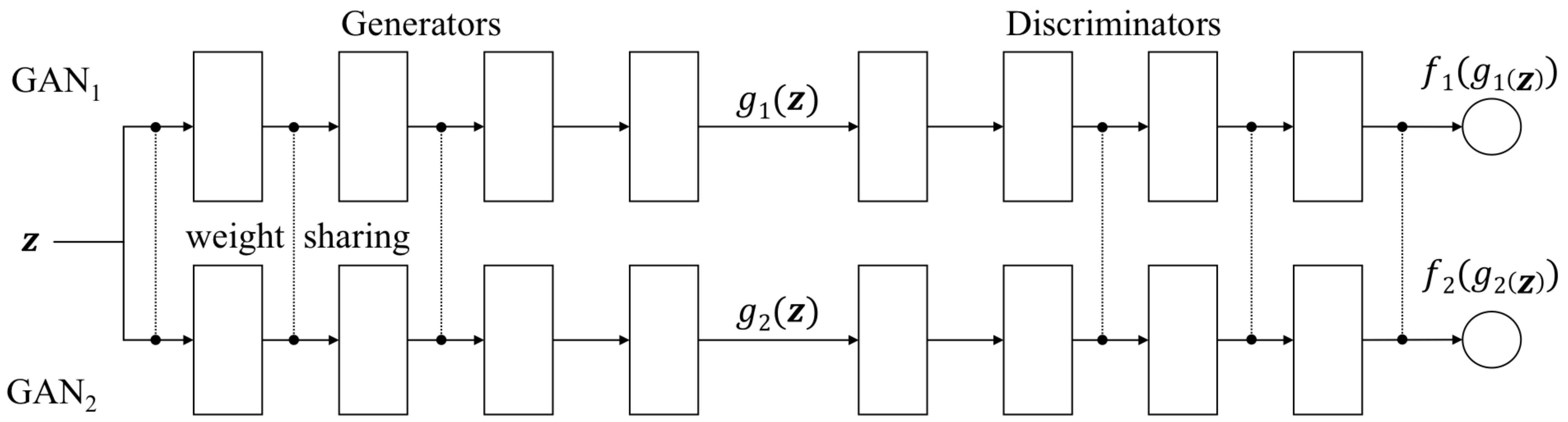
ADDA


Pre-training
source images
+ labels
Source
CNN
Classifier
class
label
Adversarial Adaptation
source images
Source
CNN
target images
Target
CNN
Discriminator
domain
label
Testing
target image
Target
CNN
Classifier
class
label

Experiments results

Method	MNIST → USPS 	USPS → MNIST 	SVHN → MNIST 
Source only	0.752 ± 0.016	0.571 ± 0.017	0.601 ± 0.011
Gradient reversal	0.771 ± 0.018	0.730 ± 0.020	0.739 [16]
Domain confusion	0.791 ± 0.005	0.665 ± 0.033	0.681 ± 0.003
CoGAN	0.912 ± 0.008	0.891 ± 0.008	did not converge
ADDA (Ours)	0.894 ± 0.002	0.901 ± 0.008	0.760 ± 0.018

Coupled GAN


Generators
Discriminators
GAN$_1$
GAN$_2$
z
weight sharing
$g_1(z)$
$g_2(z)$
$f_1(g_{1(z)})$
$f_2(g_{2(z)})$

Sources

Статья на хабре:

https://m.habr.com/company/otus/blog/358946/

Awesome domain adaptation:

https://github.com/zhaoxin94/awsome-domain-adaptation

Deep Visual Domain Adaptation: A Survey

https://arxiv.org/abs/1802.03601v4

Unsupervised domain adaptation by backpropogation

https://arxiv.org/abs/1409.7495

ADDA:

https://arxiv.org/abs/1702.05464

Coupled GANS:

https://arxiv.org/abs/1606.07536