2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-1179-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/EMBC46164.2021.9631004

Cross-Phase Adversarial Domain Adaptation for Deep Disease-free Survival Prediction with Gastric Cancer CT Images

Siwen Wang[1,2], Di Dong[1,2], Hailin Li[3], Caizhen Feng[4], Yi Wang[4], and Jie Tian[1,3], *IEEE Fellow*

***Abstract*—Predicting gastric cancer disease-free survival (DFS) and identifying patients probably with high risk are imperative for more appropriate clinical treatment plans. Compared with CT-based radiomics researches adopting linear Cox proportional hazards models, deep neural networks can perform nonlinear transformations and investigate complex associations of image features with prognosis. Exploring shared information between post-contrast CT (with better visual enhancement) and pre-contrast CT (with few side effects and contraindications) is another challenge. In this work, a cross-phase adversarial domain adaptation (CPADA) framework is proposed to adapt a deep DFS prediction network (DDFS-Net) from arterial phase to pre-contrast phase. The DDFS-Net is designed for feature learning and trained by optimizing the average negative log function of Cox partial likelihood. The CPADA maps the feature space of pre-contrast phase (target) to arterial phase (source) in an adversarial manner by measuring Wasserstein distance. The proposed methods are evaluated on a dataset of 249 gastric cancer patients by concordance index, receiver operating characteristic curves, and Kaplan-Meier survival curves. The results demonstrate that our DDFS-Net outperforms linear survival analysis methods, and the CPADA works better than supervised learning and direct transfer schemes.**

***Clinical Relevance*—This work enables preoperative DFS prediction and risk stratification in gastric cancer. It is feasible and effective to infer a patient's risk of failure given a pre-contrast CT image by DDFS-Net adapted by CPADA.**

I. Introduction

Gastric cancer (GC) ranks third among the leading causes of cancer-related death across the globe [1]. Many patients have recurrent diseases after resection with curative intent [2]. It is hence imperative to predict disease-free survival (DFS) preoperatively and identify patients probably with high risk for more suitable treatment plans.

Computed tomography (CT) is now widely applied in capturing information about recurrence and death for GC. Previous CT-based radiomics researches that quantify tumor heterogeneity and mine high-throughput image features have revealed success in evaluating prognosis for GC patients [3-5]. However, there still remain two points that need improving. For one, deep learning has shown superior performance in that deep neural networks (DNN) made up of stacked layers are capable of tackling nonlinear transformations and extracting image features that may have complex interactions with various medical tasks. For another, the Cox proportional hazards (CPH) model is usually adopted as the standard tool in survival analysis [6]. But the CPH model examines the joint effects of potential features on event occurrence risk under the assumption that a patient's risk of failure is a linear combination of these features, which may be too simplistic. Thus, the above considerations motivate us to apply DNN to learn high-order features and model survival data nonlinearly.

Given the inputs of DNN, CT images have pre-contrast phase and post-contrast phase according to whether the contrast medium is injected. Different phases represent multiple records from the same patient but with visual discrepancies (Fig. 1). Post-contrast CT is preferable because it can visualize tumor vascularity and reveal tumor heterogeneity. For example, the mucosa at the lesion can be presented as a focal enhanced line in arterial phase [7]. Pre-contrast CT, though with ordinary visual appearances, can reduce the side effects and contraindications brought by contrast medium. Thus, transferring knowledge from post-contrast to pre-contrast phase can explore more effective feature representation while mitigating side effects and contraindications. However, directly transferring the DNN trained on post-contrast CT to pre-contrast CT may result in performance degradation caused by domain shift [8], which calls for domain adaptation methods [9, 10]. Domain adaptation can successfully cope with the source and target domains by feature alignment, and adversarial learning excels in minimizing feature distances and mapping feature distributions [11, 12]. But so far, there is a lack of research on cross-phase adversarial domain adaptation.

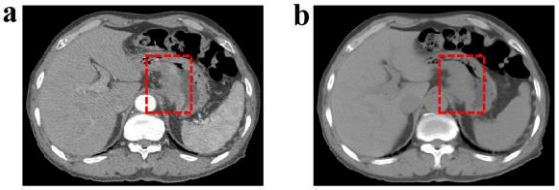

Fig. 1. Examples of a) arterial phase CT and b) pre-contrast phase CT. The tumors are framed by red boxes.

In this work, we propose a cross-phase adversarial domain adaptation (CPADA) framework to adapt a deep DFS prediction network for GC from arterial to pre-contrast phase CT. Our main contributions are as follows. First, we train a nonlinear DNN for DFS prediction by optimizing the average negative log function of Cox partial likelihood, and achieve better performance than survival analysis methods focusing on

This work was supported by National Key R&D Program of China (2017YFC1309100, 2017YFA0205200), National Natural Science Foundation of China (82022036, 91959130, 81971776, 81771924, 81930053, 81527805), Beijing Natural Science Foundation (L182061), and Youth Innovation Promotion Association CAS (2017175). The corresponding authors are Yi Wang (wangyi@pkuph.edu.cn) and Jie Tian (tian@ieee.org).

[1] CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
[2] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, 100049, China
[3] Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University, Beijing, 100191, China
[4] Department of Radiology, Peking University People's Hospital, Beijing, 100044, China

linear relationships. Second, the proposed CPADA framework maps the pre-contrast phase (target) feature space to arterial phase (source) in an adversarial manner, wherein Wasserstein distance with gradient penalty is measured to learn more indiscriminative features between domains. Third, we evaluate the stratification ability of CPADA-predicted risk values, verifying the prognostic efficacy of CPADA in providing personalized treatment plans.

II. Methodology

A. Pre-training Deep DFS Prediction Network on Source Domain

To extract high-order features related to DFS and achieve good prognostic performance, we first pre-train a deep DFS prediction network (DDFS-Net) based on arterial phase CT images X from source domain and complete survival data (DFS time T and DFS event E).

In survival analysis, the hazard function has two components: a baseline hazard and a risk item $r(x) = e^{h(x)}$ (representing the effects of potential features on baseline hazard), where $h(x)$ is a log-risk function [13]. The standard CPH model generally estimates $h(x)$ by a linear function $\hat{h}_\beta(x) = \beta^T x$ and tunes β to optimize the Cox partial likelihood:

$$L(\beta) = \prod_{E_i=1} \frac{exp\left(\hat{h}_\beta(x_i)\right)}{\sum_{j\in\mathcal{R}(T_i)} exp\left(\hat{h}_\beta(x_j)\right)}, \quad (1)$$

where $L(\beta)$ represents the probability of patient i having an event occurrence ($E = 1$) at event time T_i, given this set of patients still at risk, and x_i can represent the image data. Inspired by DeepSurv [14], we design a DDFS-Net to provide a nonlinear analysis by replacing the linear function $\hat{h}_\beta(x)$ with the network prediction $\hat{h}_\theta(x)$, where nonlinear effects of image features on the Cox partial likelihood are parameterized by the network weights θ.

Fig. 2a illustrates the architecture of DDFS-Net. We utilize the 2D-fashion deep residual network (ResNet-10) as the backbone [15] and modify the structure after the last residual block. Specially, we introduce a local average pooling (LAP) module to reduce the impact of feature noise, as motivated by the local max pooling in [16]. The LAP module works by partitioning the feature maps after the last residual block to two parts horizontally and performing average pooling on each part separately. Then, we concatenate the outputs of the two parts as the final feature representation, followed by two fully connected layers and SoftMax. The final output of the DDFS-Net represents the estimate of a patient's risk of failure.

The DDFS-Net for source domain F_S is trained and optimized by minimizing the average negative log function of Cox partial likelihood with regularization:

$$L(\theta) = \frac{-1}{N_{E=1}} \sum_{E_i=1} \left(\hat{h}_\theta(x_i) - log \sum_{j\in\mathcal{R}(T_i)} e^{\hat{h}_\theta(x_j)}\right) + \lambda\|\theta\|_2^2 , \quad (2)$$

where $N_{E=1}$ is the number of patients with event occurrence and λ is the parameter for ℓ_2 regularization. We adopt Adam optimizer with an initial learning rate of 0.0001 and a weight decay rate of 0.00001. We set the batch size to 32, and the training process is run for 200 epochs in Python (version 3.7.4) using PyTorch (version 1.7.0) on an NVIDIA GeForce RTX 2080Ti GPU.

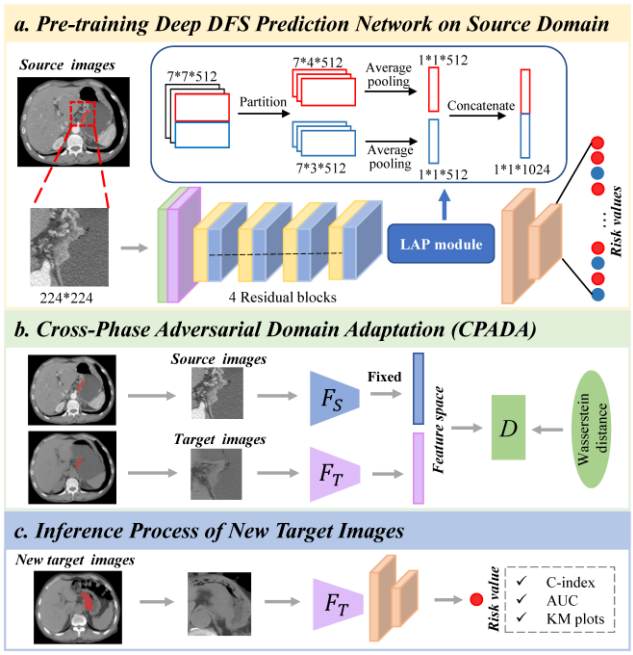

Fig. 2. The proposed cross-phase adversarial domain adaptation framework, consisting of three steps: a) pre-training deep DFS prediction network on source domain, b) cross-phase adversarial domain adaptation (CPADA), and c) inference process of new target images. KM, Kaplan-Meier.

B. Cross-Phase Adversarial Domain Adaptation (CPADA)

We then fix the feature extractor part of the pre-trained F_S and learn the DDFS-Net for target domain F_T by mapping pre-contrast phase feature space from target domain to arterial phase from source domain adversarially. Similar to the concept behind Wasserstein generative adversarial networks [17], we play a two-player minimax game (Fig. 2b). A discriminator D is designed to maximize the Wasserstein distance between feature spaces obtained by F_T and pre-trained F_S, while the DDFS-Net for target domain F_T tries to fool the discriminator D to make mistakes, *i.e.* minimizing the Wasserstein distance. The Wasserstein distance is continuous and differentiable, and thus can provide useful gradients for optimization, which ensures the convergence of CPADA.

In this case, we suppose the discriminator D is a 1-Lipschitz function, and D is optimized by maximizing the Wasserstein distance

$$L_{wd}(D) = \mathbb{E}_{X_S\sim p_S} D\left(F_S(X_S)\right) - \mathbb{E}_{X_T\sim p_T} D\left(F_T(X_T)\right), \quad (3)$$

while F_T is trained by optimizing

$$L_{wd}(F_T) = -\mathbb{E}_{X_T\sim p_T} D\left(F_T(X_T)\right). \quad (4)$$

To fulfill the Lipschitz constraint, we add a gradient penalty for discriminator D [18]:

$$L_{gp}(D) = \left(\left\|\nabla_{\hat{X}} D\left(\hat{X}\right)\right\|_2 - 1\right)^2, \quad (5)$$

where $\hat{X}$ is a set of features containing source features, target features, and random points along the straight line between source and target feature pairs. Thus, the discriminator D can be overall optimized by

$$max\ L_{wd}(D) - \alpha L_{gp}(D), \quad (6)$$

where α is an empirically set balancing coefficient.

Specifically, the discriminator D produces real (source) / fake (target) predictions via three fully connected layers followed by Leaky ReLU. During adversarial training, we

adopt Adam optimizer for F_T and stochastic gradient descent (SGD) optimizer for D (learning rate 0.001). The parameter α is practically set to 0.8.

C. Inference Process of New Target Images

In the testing stage, we realize the inference process of applying the DDFS-Net adapted by CPADA on a new pre-contrast CT image from target domain (Fig. 2c). When inputting a pre-contrast CT image of a GC patient, the adapted DDFS-Net F_T can extract features more indiscriminative between arterial phase and pre-contrast phase and predict risk values for patients.

Models are evaluated using concordance index (C-index) and area under the time-dependent receiver operating characteristic (ROC) curve (AUC) that measures DFS prediction performance at 1, 2, and 3 years. Further, using the median of patients' risk values as the cutoff, patients can be stratified into high-risk and low-risk groups of developing diseases. We plot Kaplan-Meier survival curves to explore significant differences between groups, which can verify the stratification ability and the prognostic efficacy of models in providing personalized treatment plans.

III. Dataset and Results

A. Dataset and Preprocessing

This work retrospectively collect 249 GC patients from Peking University People's Hospital (Table I). The Institution's Ethical Review Board approved all experimental procedures involving human subjects. Pre-contrast and arterial phase CT images as well as complete DFS information are collected. The DFS event is defined as local recurrence, metachronous metastatic disease, or death caused by GC. If an DFS event occurrence is observed ($E = 1$), the DFS time T represents the time elapse between preoperative CT scan and the event occurrence time. If an DFS event occurrence is not observed ($E = 0$), the DFS time T ranges from preoperative CT scan to the last follow-up visit. We randomly allocate the patients to a training set (166 patients) and a testing set (83 patients) at a 2:1 ratio. More details about patient enrollment can be found in our previous work [5].

TABLE I. Clinical Information of Gastric Cancer Patients

Clinical information		Training (n=166)	Testing (n=83)
Age, mean ± SD, years		64.1±12.3	61.3±12.9
Sex, No. (%)			
	Male	122 (73.5)	60 (72.3)
	Female	44 (26.5)	23 (27.7)
DFS time, median (IQR),		25.5 (12.0-46.0)	22.0 (12.0-45.0)
DFS event, No. (%)			
	$E = 1$	58 (34.9)	33 (39.8)
	$E = 0$	108 (65.1)	50 (60.2)

NOTE. SD, standard deviation; IQR, interquartile range.

The tumor regions of interest are manually segmented on the slice with the largest tumor area for both pre-contrast and arterial phase CT images by two radiologists upon consensus. We crop the images by corresponding axis-aligned minimum bounding boxes base on radiologists' segmentations and dilate the cropped patches within an adaptive small range. Meanwhile, we adjust the window width to 500HU and window level to 30HU to ensure appropriate and unified CT image intensities. All the patches are then resized to 224×224.

B. Effectiveness of Deep DFS Prediction Network

When comparing linear and nonlinear methods for DFS prediction as presented in Table II, we adopt the same testing set as that in [5] for a fair comparison. For linear methods on pre-contrast and arterial phase CT images, we also follow the same procedures as in [5]: extracting and selecting features by radiomics and fitting features by linear CPH model.

TABLE II. Comparison of Linear and Nonlinear Methods for DFS Prediction on Testing Set

Configurations	C-index (95% CI)	Time-dependent AUC		
		1-year	2-year	3-year
Pre-contrast (Linear)	0.596 (0.509-0.683)	0.649	0.641	0.695
Pre-contrast (Nonlinear)	0.608 (0.509-0.707)	0.659	0.608	0.579
Arterial (Linear)	0.616 (0.526-0.705)	0.628	0.635	0.691
Arterial (Nonlinear)	0.707 (0.594-0.819)	0.711	0.706	0.658

NOTE. "Linear" denotes feature extraction and selection by radiomics and feature fitting by linear Cox proportional hazards model. "Nonlinear" denotes our nonlinear DDFS-Net. CI, confidence interval.

The nonlinear DDFS-Net effectively improves the prognostic performance compared to linear methods as indicated by C-indices of 0.707 *vs.* 0.616 for arterial phase CT images and 0.608 *vs.* 0.596 for pre-contrast images. Specifically, the time-dependent AUCs for DDFS-Net on arterial phase are largely promoted at 1 and 2 years compared to linear methods. Also, the DDFS-Net can separate patients with more significant log-rank P values (0.0005 *vs.* 0.006 and 0.004 *vs.* 0.1, Fig. 3), indicating the superior risk stratification ability of nonlinear methods. Besides, we can observe that pre-contrast images yield a poor prognostic performance, no matter which method is adopted. This is probably because pre-contrast images suffer from information loss, while arterial phase may provide more useful and detailed information related to DFS.

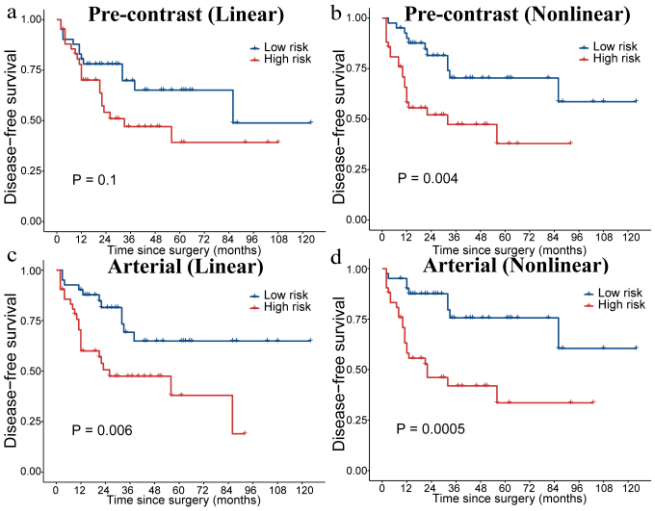

Fig. 3. Survival analysis for linear and nonlinear methods. "Linear" denotes feature extraction and selection by radiomics and feature fitting by linear Cox proportional hazards model. "Nonlinear" denotes our nonlinear DDFS-Net.

TABLE III. COMPARISON OF DIFFERENT METHODS ON TARGET DOMAIN USING PRE-CONTRAST CT

Methods	Configurations		C-index (95% CI)	Time-dependent AUC		
	Training	Testing		1-year	2-year	3-year
Supervised Learning	$Target_{train}$	$Target_{test}$	0.608 (0.509-0.707)	0.659	0.608	0.579
Direct Transfer	$Source_{train}$	$Target_{test}$	0.514 (0.409-0.618)	0.529	0.521	0.519
CPADA	$F_T(Target_{train})$	$Target_{test}$	0.649 (0.558-0.740)	0.636	0.630	0.587

NOTE. "Supervised Learning" denotes training and testing DDFS-Net on pre-contrast CT images from the target domain. "Direct Transfer" means directly applying the source domain pre-trained DDFS-Net on the target domain. "CPADA" represents testing the CPADA adapted DDFS-Net on the target domain. $Source_{train}$, $Target_{train}$, and $Target_{test}$ denote the training set of source domain, the training set and testing set of target domain, respectively.

C. Comparison with Supervised Learning and Direct Transfer

Briefly, the comparison results of the proposed CPADA framework with different methods on target domain are summarized in Table III. The supervised learning scheme denotes training and testing on target domain. The direct transfer scheme directly applies the source domain pre-trained DDFS-Net F_S on target domain.

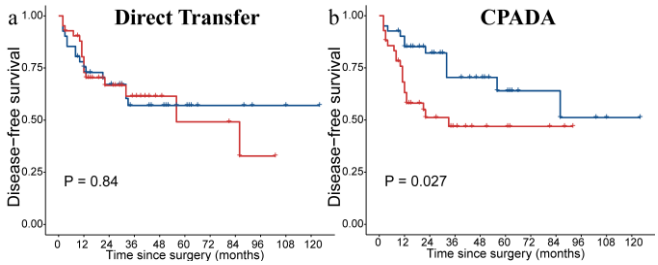

Fig. 4. Risk stratification ability comparison for direct transfer scheme and CPADA framework.

We can clearly see a large drop in C-index and AUCs when using the direct transfer scheme, the reason behind which should be the different feature space distributions between pre-contrast and post-contrast image phases [7]. This is further verified in Fig. 4a that the directly transferred model even fails in risk stratification (P = 0.84). Using the CPADA framework, however, the prognostic performance gains improvements of 6.7% and 26.3% in C-indices, respectively, compared to supervised learning and direct transfer. The adapted DDFS-Net extracts features more indiscriminative between different phases. Through the CPADA framework, DFS prediction on a new pre-contrast image can be done more safely, conveniently, and accurately.

IV. CONCLUSION

This work proposes a CPADA framework to adapt the deep DFS prediction network from arterial phase CT images to pre-contrast phase, which improves the prognostic performance and enables more appropriate and safer treatment decision making for GC patients. Experimental results show that our nonlinear DDFS-Net effectively improves the prognostic performance compared to linear methods, and comparison results indicate that the CPADA framework outperforms supervised learning and direct transfer schemes in DFS prediction and risk stratification. In the future, a more practical scenario with multiple source domains, *i.e.*, arterial, venous, and delayed phases can be considered and explored. Also, models will be further validated on external datasets.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA : a cancer journal for clinicians, vol. 68, no. 6, pp. 394-424, 2018.

[2] E. Van Cutsem, X. Sagaert, B. Topal, K. Haustermans, and H. Prenen, "Gastric cancer," The Lancet, vol. 388, no. 10060, pp. 2654-2664, 2016.

[3] D. Dong et al., "Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer," Annals of Oncology, vol. 30, no. 3, pp. 431-438, 2019.

[4] W. Li et al., "Prognostic value of computed tomography radiomics features in patients with gastric cancer following curative resection," European radiology, vol. 29, no. 6, pp. 3079-3089, 2019.

[5] S. Wang et al., "Preoperative computed tomography-guided disease-free survival prediction in gastric cancer: a multicenter radiomics study," Medical Physics, vol. 47, no. 10, pp. 4862-4871, 2020.

[6] D. R. Cox, "Regression models and life-tables," Journal of the Royal Statistical Society: Series B (Methodological), vol. 34, no. 2, pp. 187-202, 1972.

[7] X.-H. Chen, K. Ren, P. Liang, Y.-R. Chai, K.-S. Chen, and J.-B. Gao, "Spectral computed tomography in advanced gastric cancer: Can iodine concentration non-invasively assess angiogenesis?," World Journal of Gastroenterology, vol. 23, no. 9, p. 1666, 2017.

[8] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in CVPR 2011, 2011, pp. 1521-1528: IEEE.

[9] G. Wilson and D. J. Cook, "A survey of unsupervised deep domain adaptation," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 11, no. 5, pp. 1-46, 2020.

[10] S. Zhao et al., "A Review of Single-Source Deep Unsupervised Visual Domain Adaptation," IEEE Transactions on Neural Networks and Learning Systems, 2020.

[11] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7167-7176.

[12] Y. Ganin et al., "Domain-adversarial training of neural networks," The journal of machine learning research, vol. 17, no. 1, pp. 2096-2030, 2016.

[13] T. M. Therneau and P. M. Grambsch, "Estimating the survival and hazard functions," in Modeling survival data: extending the cox model: Springer, 2000, pp. 7-37.

[14] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network," BMC medical research methodology, vol. 18, no. 1, pp. 1-12, 2018.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[16] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, and J. Jiao, "Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 994-1003.

[17] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in International conference on machine learning, 2017, pp. 214-223: PMLR.

[18] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," arXiv preprint arXiv:1704.00028, 2017.