Check for updates

Received: 13 December 2021 | Revised: 21 June 2022 | Accepted: 12 July 2022

IET Intelligent Transport Systems

DOI: 10.1049/itr2.12251

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Multiple auxiliary classifiers GAN for controllable image generation: Application to license plate recognition

Nadiya Shvai[1] | Abul Hasnat[1] | Amir Nakib[2]

[1]Cyclope.ai, VINCI Autoroutes, Paris, France

[2]University Paris Est Créteil, Laboratoire LISSI, Vitry sur Seine, France

Correspondence
Amir Nakib, University Paris Est Créteil, Laboratoire LISSI, Vitry sur Seine, France.
Email: nakib@u-pec.fr

Funding information
vinci autoroutes and cyclope.ai

Abstract

One of the main challenges in developing machine learning (ML) applications is the lack of labeled and balanced datasets. In the literature, different techniques tackle this problem via augmentation, rendering, and over-sampling. Still, these methods produce datasets that appear less natural, exhibit poor balance, and have less variation. One potential solution is to leverage the Generative Adversarial Network (GAN) which achieves remarkable results in the generation of high-fidelity natural images. However, expanding the ability of GANs' to control generated image attributes with supervisory information remains a challenge. This research aims to propose an efficient method to generate high-fidelity natural images with total control of its main attributes. Therefore, this paper proposes a novel Multiple Auxiliary Classifiers GAN (MAC-GAN) framework based on Auxiliary Classifier GAN (AC-GAN), multi-conditioning, Wasserstein distance, gradient penalty, and dynamic loss. It is therefore presented as an efficient solution for highly controllable image synthesis red that allows to enrich and re-balance datasets beyond data augmentation. Furthermore, the effectiveness of MAC-GAN images on a target ML application called Automatic License Plate Recognition (ALPR) under limited resource constraints is probed. The improvement achieved is over 5% accuracy, which is mainly due to the ability of the MAC-GAN to create a balanced dataset with controllable synthesis and produce multiple (different) images with the same attributes, thus increasing the variation of the dataset in a more elaborate way than data augmentation techniques.

1 | INTRODUCTION

GAN [1] based natural image generation is consistently reaching new milestones. Indeed, since its introduction, the potentials of GAN have been widely explored, and today it is often very challenging for a human to distinguish a GAN synthesized image (e.g. face) from a real one [2, 3]. This highlights GAN's ability to become one of the most useful algorithms for the development and evaluation of different ML methods. However, to gain such a level of usability, ML practitioners should have enough control over GAN synthesis to represent different scene attributes. To fulfill this requirement, we proposed a novel method, called MAC-GAN, that provided good results in generating controllable and diverse images.

The basic idea of GAN is to simultaneously train two networks (generator and discriminator) based on a competitive learning strategy [1]. While the goal of the generator is to model the latent space of input data distribution and synthesize realistic samples, the discriminator aims at distinguishing the synthesized samples from the real ones. One of the main purposes of training GANs is to achieve a perfect generator that provides realistic samples. Odena et al. [4] proposed the AC-GAN strategy to generate high-quality natural images with a single input/condition based on a pre-defined set of class names. This research aims to extend AC-GAN's flexibility to generate images with multiple conditions based on different (e.g. class name, background) attributes. To the best of our knowledge, multi-conditioning with AC-GAN has not yet been explored for image synthesis.

We develop MAC-GAN by leveraging several frameworks and optimization strategies. Its architecture exploits AC-GAN [4] with multi-conditioning to achieve controlla-

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2022 The Authors. *IET Intelligent Transport Systems* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

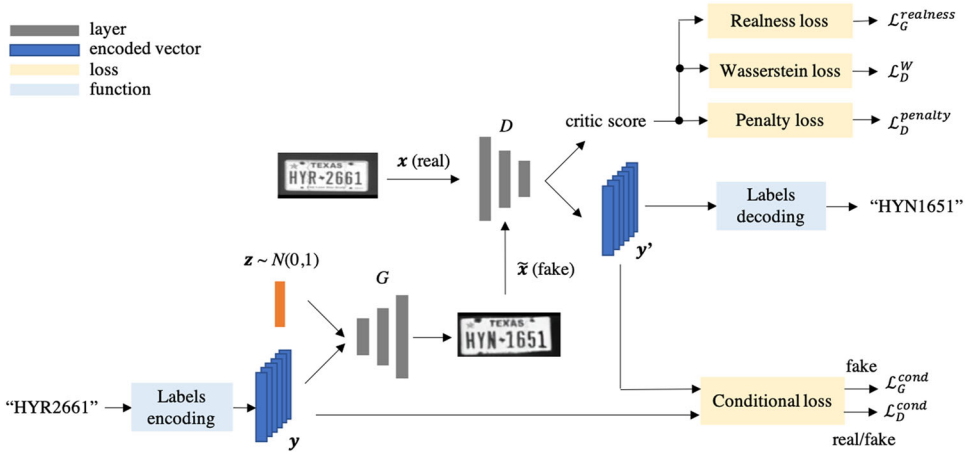


FIGURE 1 Architecture of the proposed MAC-GAN framework. MAC-GAN accepts as input a condition vector $\boldsymbol{y}$ (in the form of one-hot encoded labels) and a noise vector $\boldsymbol{z}$. In the case of license plate generation, we consider the condition vector $\boldsymbol{y}$ to be the license plate text. The generator component G synthesises image based on the input, while the critic component D assesses image realness and its correspondence to the input condition vector. The proposed framework allows one to gain explicit control over desirable attributes of the generated output while maintaining image quality and diversity

bility over different image attributes. Moreover, it adopts the Progressive Growing of GANs (PG-GAN) [5] strategy to generate high-resolution images. It incorporates Wasserstein distance and gradient penalty to ensure the generation of better quality images. Finally, MAC-GAN optimizes a dynamic loss during training, that consists of a dynamic conditioning weighting scheme for the coefficients of its multi-objectives formulation.

This work aims at providing complete user control for GAN based image synthesis to overcome the data scarcity challenges. More generally, for a given ML use case, the objective is to increase samples variability and to balance samples distribution by generating images with selective attributes. We consider the real-world ALPR task as the target ML application to evaluate MAC-GAN's effectiveness. Indeed, synthesizing the license plate (LP) text images is an ideal task to evaluate MAC-GAN's controllability, as it provides extensive possibilities for the input controls. For example, an LP image of length 7 with 36 possible characters at each location can be generated from 36^7 possible choices as input. This research does not consider the other ML tasks and datasets, because they are less suitable for evaluating image synthesis based on multiple labels. Indeed, images and labels from them have one of the following restrictions: (a) single attribute/image class [6, 7]; (b) multiple binary attributes [8] and (c) unstructured inputs [9, 10].

The rapid progress of deep convolutional neural networks (CNNs) based algorithms allows the recent ALPR systems to achieve very high accuracy [11–13]. However, these methods require a large number of labelled images for training. A large training dataset ensures high sample variations and balance on character distributions, imaging conditions and the licence plate (LP) structures [12]. Indeed, the lack of variations harms both training and evaluation [13–15]. Furthermore, we observe that even a large LP dataset might suffer from imbalanced character joint distribution as well as low appearance variation for unique LP text images. On the other hand, currently, it is very challenging to enhance the ALPR datasets due to the restrictions on private data storage, such as the *General Data Protection Regulation* (GDPR) [16], which prevents data storage for more than 30 days. Therefore, the problems of data scarcity and lack of variations are further exacerbated. An obvious solution is to develop a framework that generates high-quality LP images with complete control of the characters. This research is strongly motivated by this requirement and explores MAC-GAN for LP generation.

Existing LP generation methods [13–15] apply image-to-image translation based GAN methods [17, 18], which exhibit several limitations concerning image quality and sample variations. The proposed MAC-GAN aims to alleviate these limitations. Experiments indicate that it successfully allows to fully control the image generation (Figure 3) and helps to achieve additional ≈5% ALPR accuracy under different resource constraints (Figure 5a). Overall, the key contributions of this research can be summarized as:

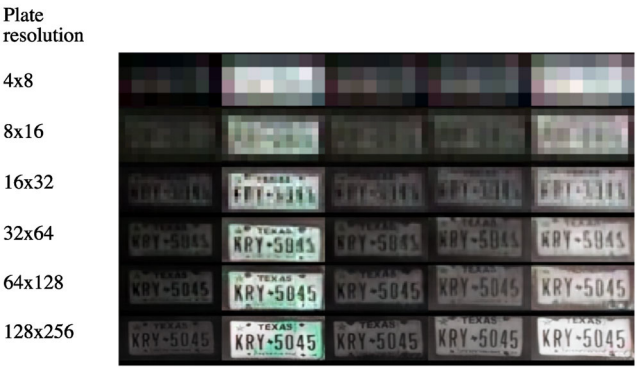


FIGURE 2 Evolution of the controlled generation along with training. Each row represents the training progress of the generator for the given input "KRY5045" when increasing the resolution (from 4x8 to 128x256) and each column corresponds to a random latent vector used for the image generation

17519578, 2023, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.12251 by Readcube (Labtiva Inc.), Wiley Online Library on [10/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

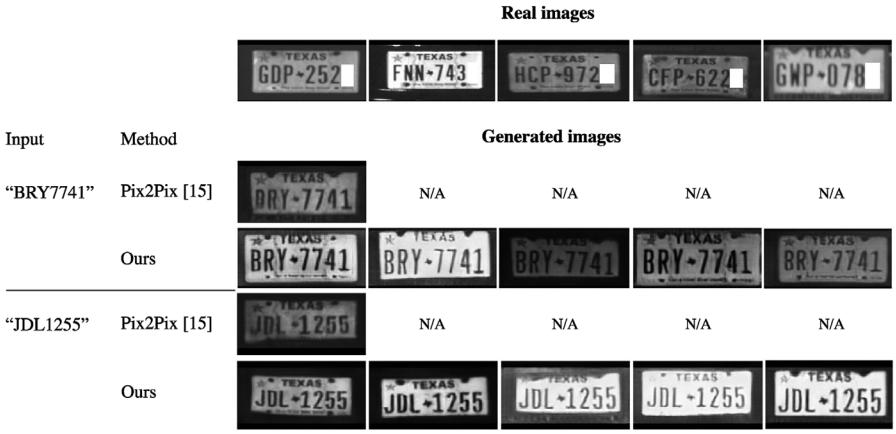


FIGURE 3 Examples of controlled images generation for 2 given inputs. [17] generates only one image for a given input. Our method generates diverse LP images for different random latent vectors feeding the generator. Real images from the dataset are provided for visual comparison, with the last character removed for privacy purposes. LP texts on the generated images are imaginary

(a) we propose a novel GAN based method to produce *high quality and fully controllable* images through *multi-label conditioning*. The ability to generate such images helps to construct an enhanced training dataset with more variability and balanced distribution of its attributes, which cannot be achieved with existing image augmentation techniques. Moreover, MAC-GAN generated images are superior to the state-of-the-art with respect to both quality and variability;
(b) we demonstrate the significant impact of *controllable image generation for the ALPR application under resource constraint*, which clearly shows the usefulness of the proposed method; and
(c) we introduce new *evaluation metrics* for the GAN methods specialized for text generation. This will help to green benchmark similar approaches with common metrics.

The rest of the paper is organized as follows: Section 2 is dedicated to related work. The proposed MAC-GAN framework is explained in Section 3. The experiments and results are described in Section 4. Finally, Section 5 provides conclusions of the study.

2 | RELATED WORK

GAN based methods demonstrate remarkable results at generating high fidelity images [2, 3, 5]. However, the challenge of fully controlling the generated images through structured data is relatively under-explored. We study the literature from this perspective and explore the possible approaches and applications.

2.1 | Conditional generation

Historically GANs appear in an unsupervised setting where the goal is to generate real and natural images. In the most basic settings , GANs use a generator to synthesize images and a discriminator to classify images as real or fake. Several approaches enhanced GANs to include supervised information.

Conditional GAN (C-GAN) [19] extended traditional GAN to generate better images. C-GAN includes conditions on both generator and discriminator based on additional information y.

Another approach is AC-GAN [4], which extended the original goal of GAN to perform two tasks: (a) classify images as real or fake and (b) predict true attribute information y. However, unlike modeling the joint distribution (as C-GAN), AC-GAN models the conditional distribution of the attribute y and image x. Inspired by this strategy, generated samples can be conditioned on classes information [20, 21], text [22], bounding box [23], pose estimation [24] etc.

In [25] a problem of topology mismatch between latent and output spaces for conditional GANs has been tackled. The proposed method encourages a bi-lipschitz mapping between the latent and the output manifolds thus improving the quality of synthetic images *w.r.t.* image diversity and realness.

Numerous complex architectures apply conditional generation and demonstrate good results for anime generation [26], brain metastases synthesis [27], 3D brain images at different stages of the Alzheimer's disease [28] or gastritis image generation [29].

However, these previous works take place in a single-label setting, where the input consists of the label vector y and random noise vector z. The discriminator produces two outputs, the probability of x being real or fake, and the estimated conditional probability $p(x|y)$. *While the use of label vector y enlightens great potential at controlling the generated output, there is no existing work, to our knowledge, on the possibility of gaining further control within a multi-label context*. Therefore, this research aims to expand single-level AC-GAN into a multi-level context to develop a fully controllable image generator framework. To demonstrate its effectiveness, we choose the licence plate (LP) text generation task to enhance the ALPR systems.

17519578, 2023, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/itr2.12251 by Readcube (Labtiva Inc.), Wiley Online Library on [10/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2 | License plate generation

In the past few years, CNN based ALPR methods [11] have achieved great success. However, they require a massive quantity of labelled data which can be costly and time-consuming to collect. Furthermore, existing systems tend to over-fit (respectively under-perform) on the over-represented (respectively under-represented) characters. To overcome these limitations, LP image synthesis methods have been proposed to generate new examples of labelled training data and hence improve the ALPR systems [13–15]. Wang et al. [14] improves the CycleGAN based model [18] that learns the mapping between the synthetic (generated by a script) and real images. Wu et al. [15] adopt a similar approach and improve the image-to-image based Pix2Pix generative architecture [17]. Although these methods demonstrated good results, they have particular limitations such as poor quality of the generated images and no variation (i.e. single image generation) for an input LP text. This research aims to overcome these limitations.

In [30] Silvano et al. propose to generate large volumes of accurate Mercosur standard LP images with template-based synthesis. They used data augmentation that mimics real-life conditions such as artificial shading and sloping. Such an approach allows to obtain images with any LP text, however it is difficult to extend as it heavily relies on a specific template and data augmentation.

2.3 | Text-to-image generation

Text-to-image generation can be considered one of the most relevant applications of *controlled image generation task*. It aims at generating realistic images that match the user given text descriptions.

Reed et al. [31] use conditional PixelCNN to generate images using the text descriptions and objects locations constraints. Zhang et al. [22, 32, 33] decompose the task into several stages from coarse to fine level. Li et al. [34] introduces a word-level discriminator and channel-wise attention-driven generator. Park et al. [35] defines a novel GAN based architecture that generates new images by synthesizing the background of an original image and a new object described by the text description. Gao et al. proposed Lightweight Dynamic Conditional GAN [36] for text-to-image synthesis where the attention block was used to capture multi-scale context. All of the above approaches generate images from unstructured data. In our case, we use multiple conditioning labels to represent the LP text structure and generate controlled images. This is the key difference with the previous work.

3 | Multiple auxiliary classifiers GAN

Given multiple conditions encoded in $\boldsymbol{y}$, we aim to synthesize an image $\tilde{\boldsymbol{x}}$, that appears realistic and respects the conditions during the generation step. To achieve this, we propose two novel components: MAC-GAN framework (Figure 1), and dynamic loss based on conditional weights (Section 3.3) for effective training.

3.1 | Architecture

Figure 1 illustrates the overall architecture of MAC-GAN, which consists of two neural networks, generator G and critic[1] D.

Generator G generates fake images $\tilde{\boldsymbol{x}} = G(\boldsymbol{z},\boldsymbol{y})$, where input vector $\boldsymbol{z}$ is noise, which is a fixed-length random vector sampled from a Gaussian distribution. Input $\boldsymbol{y} = (y_1, \dots, y_K)$ is the condition vector, where every component y_j corresponds to one independent discrete condition with a fixed number of values. Critic D has two distinct types of outputs, raw critic score D^r and a set of K probability vectors D^j, $D(\tilde{\boldsymbol{x}}) = (D^r(\tilde{\boldsymbol{x}}), D^1(\tilde{\boldsymbol{x}}), \dots, D^K(\tilde{\boldsymbol{x}}))$. The raw critic score D^r estimates the Wasserstein-1 distance between the distributions of the real $\boldsymbol{x}$ and generated $\tilde{\boldsymbol{x}}$ images [38, 39]. A set of softmax layers in the network D estimates the probability vectors of D^j. Each vector of D^j associates an input condition y_j, which is encoded via one-hot encoding. We use cross-entropy to compare D^j with y_j. Thus, critic D exercises control over the image realness and image compliance to the conditions encoded via $\boldsymbol{y}$.

3.2 | Objective functions

We train the generator G and critic D alternatively by minimizing both the generator loss $\mathcal{L}_G$ and critic loss $\mathcal{L}_D$. As the loss function , we use the Wasserstein GAN loss with gradient and epsilon penalty (WGAN-GEP) [38, 39] with additional terms of conditional loss. These conditional loss terms enable control over the conditions satisfaction, and each of them is independent.

3.2.1 | Generator objective

the generator loss $\mathcal{L}_G$ comprises a realness and a conditional component. The conditional loss represents a weighted sum of losses per condition with weights $\lambda_j^{cond,t}$, see Equation (1). Weights $\lambda_j^{cond,t}$ depend on epoch t, Section 3.3.1 provides further details.

$$\mathcal{L}_G^t = \mathcal{L}_G^{\text{realness}} + \mathcal{L}_G^{\text{cond}} = \mathcal{L}_G^{\text{realness}} + \sum_{j=1}^{K} \lambda_j^{cond,t} \mathcal{L}_{G,j}^{\text{cond}}. \quad (1)$$

The realness loss improves the generation quality and makes the synthetic images appear as real; the conditional loss penalizes the generated images whose labels prediction do not match the input labels. We use the critic's output $D(\tilde{\boldsymbol{x}}) =$

[1] The critic D is a substitute of classical GAN discriminator, for example, D in WGAN [37] formulation. One of the main differences between the notions of discriminator and critic is the range of output values: discriminator determines the realness probability, which lies between 0 and 1, and critic returns a critic score, which is not bound a priori.

17519578, 2023, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ntw2.12251 by Readcube (Labtiva Inc.), Wiley Online Library on [10/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$(D^r(\tilde{\boldsymbol{x}}), D^1(\tilde{\boldsymbol{x}}), \ldots, D^K(\tilde{\boldsymbol{x}}))$ to compute both of these losses. Following [37], we define the realness loss as:

$$\begin{aligned}\mathcal{L}_G^{\text{realness}} &= \mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}}\sim\mathbb{P}_g|\boldsymbol{y},\boldsymbol{y}\sim\mathbb{P}_r^{\boldsymbol{y}}}\left[-D^r(\tilde{\boldsymbol{x}})\right] \\ &= \mathop{\mathbb{E}}_{\boldsymbol{z}\sim\mathbb{P}_z,\boldsymbol{y}\sim\mathbb{P}_r^{\boldsymbol{y}}}\left[-D^r(G(\boldsymbol{z},\boldsymbol{y}))\right],\end{aligned} \tag{2}$$

where $\mathbb{P}_r$ is real data distribution, $\mathbb{P}_g$ is the distribution of the output from $G(\boldsymbol{z},\boldsymbol{y})$, and $\mathbb{P}_z$ is the distribution of noise.

We define the conditional loss for label y_j through cross-entropy on the corresponding output of critic D, that is $D^j(\tilde{\boldsymbol{x}})$. In order to implement this, we represent label y_j as a one-hot-encoded vector, which is compared to the probability vector output $D^j(\tilde{\boldsymbol{x}})$.

$$\begin{aligned}\mathcal{L}_{G,j}^{\text{cond}} &= \mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}}\sim\mathbb{P}_g|\boldsymbol{y},\boldsymbol{y}\sim\mathbb{P}_r^{\boldsymbol{y}}}\mathcal{L}^{cross}(D^j(\tilde{\boldsymbol{x}}),y_j) \\ &= \mathop{\mathbb{E}}_{\boldsymbol{z}\sim\mathbb{P}_z,\boldsymbol{y}\sim\mathbb{P}_r^{\boldsymbol{y}}}\mathcal{L}^{cross}(D^j(G(\boldsymbol{z},\boldsymbol{y})),y_j),\end{aligned} \tag{3}$$

where $\mathcal{L}^{cross}$ is the cross-entropy loss function.

Finally, we obtain an explicit form of the generator loss by combining Equations (1), (2), and (3) as:

$$\begin{aligned}\mathcal{L}_G^t = \mathop{\mathbb{E}}_{\boldsymbol{z}\sim\mathbb{P}_z,\boldsymbol{y}\sim\mathbb{P}_r^{\boldsymbol{y}}}&[-D^r(G(\boldsymbol{z},\boldsymbol{y}))+ \\ &\sum_{j=1}^{K}\lambda_j^{cond,t}\mathcal{L}^{cross}(D^j(G(\boldsymbol{z},\boldsymbol{y})),y_j)].\end{aligned} \tag{4}$$

3.2.2 | Critic objective

Following previous work on improved critic losses [4, 37, 40], critic objective $\mathcal{L}_D$ comprises the Wasserstein loss and several penalty terms such as gradient penalty, epsilon penalty , and extra conditional term. This term penalizes the images whose labels predictions do not match the input labels/conditions.

$$\begin{aligned}\mathcal{L}_D^t &= \mathcal{L}_D^W + \mathcal{L}_D^{cond} + \mathcal{L}_D^{penalty} \\ &= \mathcal{L}_D^W + \sum_{j=1}^{K}\lambda_j^t\mathcal{L}_{D,j}^{\text{cond}} + \left(\lambda^{grad}\mathcal{L}_D^{grad} + \varepsilon\mathcal{L}_D^{\varepsilon}\right).\end{aligned} \tag{5}$$

The Wasserstein loss expresses the difference between critic scores on real and generated images. It is an approximation of the Wasserstein distance among the real and generated images distributions.

$$\mathcal{L}_D^W = \mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}}\sim\mathbb{P}_g}[D^r(\tilde{\boldsymbol{x}})] - \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathbb{P}_r}[D^r(\boldsymbol{x})]. \tag{6}$$

The conditional loss allows the critic to predict the correctness of label y_j in both real and generated images. Similar to the generator's conditional loss, we define it using the cross-entropy function. We use the same weight coefficients for the conditional loss terms of the generator and critic.

$$\begin{aligned}\mathcal{L}_{D,j}^{\text{cond}} = &\mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}}\sim\mathbb{P}_g|\boldsymbol{y},\boldsymbol{y}\sim\mathbb{P}_r^{\boldsymbol{y}}}\mathcal{L}^{cross}(D^j(\tilde{\boldsymbol{x}}),y_j)+ \\ &\mathop{\mathbb{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mathbb{P}_r^{(x,y)}}\mathcal{L}^{cross}(D^j(\boldsymbol{x}),y_j) \\ = &\mathop{\mathbb{E}}_{\boldsymbol{z}\sim\mathbb{P}_z,\boldsymbol{y}\sim\mathbb{P}_r^{\boldsymbol{y}}}\mathcal{L}^{cross}(D^j(G(\boldsymbol{z},\boldsymbol{y})),y_j)+ \\ &\mathop{\mathbb{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mathbb{P}_r^{(x,y)}}\mathcal{L}^{cross}(D^j(\boldsymbol{x}),y_j).\end{aligned} \tag{7}$$

To stabilize the training procedure [40], we penalize the critic loss by the norm of the gradient of the critic scores with respect to its input. The gradient penalty constrains the 1-Lipschitz condition on the raw critic score D^r. This condition is necessary to use the Wasserstein-1 distance representation obtained through Kantorovitch-Rubinstein duality [37, 41]. Moreover, we use an epsilon penalty term to avoid loss drifting away from zero [38].

$$\begin{aligned}\mathcal{L}_D^{penalty} = &\underbrace{\lambda^{grad}\mathop{\mathbb{E}}_{\hat{\boldsymbol{x}}\sim\mathbb{P}_{\hat{x}}}\left[\left(\left\|\nabla_{\hat{x}}D^r(\hat{\boldsymbol{x}})\right\|_2-1\right)^2\right]}_{\text{gradient penalty term}}+ \\ &\underbrace{\varepsilon\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathbb{P}_r}D^r(\boldsymbol{x})^2}_{\text{epsilon penalty term}}.\end{aligned} \tag{8}$$

Here distributions $\mathbb{P}_r$ and $\mathbb{P}_g$ constructs $\mathbb{P}_{\hat{x}}$ based on a random (uniform) convex combination of the real $\boldsymbol{x}$ and generated sample $\tilde{\boldsymbol{x}}$ to construct $\hat{\boldsymbol{x}}$.

3.3 | Learning strategy

3.3.1 | Dynamic loss

As formulated in Section 3.2, the generator and critic losses include the conditional components. The weights that one chooses to apply on these terms impact the training as well as the quality of the results. To tackle this problem, we dynamically modify the combination of the different criteria during the training process. We start from no constraints on the labels correspondence. Next, we apply a stepwise increase of the weight on this conditional loss. This ensures more importance to the label correspondence (for generator) and label recognition (for critic) at the subsequent epochs of the training.

Experimentally we observe that the use of conditioning weights $\lambda_k^{cond,t}$ of Equation (9) outperforms the results obtained with a fixed conditioning value. The consequence of applying

17519578, 2023, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.12251 by Readcube (Labtiva Inc.), Wiley Online Library on [10/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

this weighting scheme is the formulation of a *dynamic loss*, which is one of the most prominent elements of MAC-GAN.

$$\lambda_k^{cond,t} = \begin{cases} 0, & t \leqslant 8 \\ 1, & t \leqslant 16 \\ 10i, & t \leqslant 8(i+2), i = 1, \dots, 7 \end{cases} . \quad (9)$$

We observe that the attributes of the generated images are similar at the beginning of the training, due to the early stages of generator's objective optimization. At this step, we set the conditional weights to 0. Once more images are available to the generator and critic, we change the weight to 1 to affect the training. Next, we apply a stepwise function to increase the weight by a factor of 10. In our experiments, we use the same schedule for all the conditions due to their complex similarity and related nature. However, in the general case, the schedules might vary to reflect the difficulty and the importance of a specific condition, and its alignment with the image generation learning phases.

3.3.2 | Data oversampling

as a part of the multi-label learning context, data might suffer from multi-label imbalance - some labels being severely under-represented compared to others. Different techniques are available to address the imbalance in multi-label datasets [32, 42–44]. These methods follow one of these approaches: data resampling, algorithm adaptation, and cost-sensitive learning. We point out that applying simple data oversampling through image augmentation techniques leads to much better results for multi-label generation as the critic learns the labels independently, and thus is robust to inter-labels correlations in the training dataset.

3.4 | Implementation

We propose a multi-conditioning framework that can adapt to existing GAN architectures and provide their discriminator the ability to learn multiple labels during adversarial training. To implement the MAC-GAN framework, we exploit the state-of-the-art Progressive Growing of GANs (PG-GANs) [5] architecture due to its excellent performance in generating high-resolution images. In PG-GANs, image resolution is progressively increased during training as new intermediate layers are added to the model. Karras et al. proves that the progressive growing techniques not only speeds-up the training, but also stabilize it. Training ensures that the labels are progressively learnt by the discriminator and the generation is oriented towards the multiple controlled inputs fed to the generator. Figure 2 illustrates the controlled generated outputs during the training progress with PG-GANs.

MAC-GAN is built upon the official TensorFlow [45] implementation of PG-GANs, from which we inherit most of the training details. In particular, we use the same generator and critic networks (with required modification of the generator's input and the critic's output), resolution-dependent mini-batch sizes, Adam hyper-parameters, activation layers, pixel-wise normalization, and exponential moving average of the generator. Both networks consist of replicated 3-layer blocks that are introduced one by one during the training phase. The detailed architectures are available in Table 3 of Appendix A.1. The input latent vector corresponds to a 512-dimensional normal random noise.

4 | EXPERIMENTS

We conduct extensive experiments on generating high-resolution natural LP images, and their impact on improving ALPR systems. To this aim, first, we collect a relevant LP image dataset. Next, we define several metrics based on [4] to evaluate the quality and variations of the generated images. Finally, we perform a competitive evaluation to demonstrate MAC-GAN's effectiveness and impact.

4.1 | Data collection

To train and evaluate, we have collected a cropped LP image dataset of 83K images which comprises 33K unique LP texts. These images were originally collected from different types (RGB and infrared) of cameras located in the free-flow tolling systems. The collected dataset ensures a large variation of the LP images with respect to image capturing conditions and appearances. We split the dataset for training and test, where the training set comprises 75K images of 30K unique LPs and the test set comprises 8K images of 3K unique LPs. For a fair evaluation, we ensure no overlap among the training and test sets, that is, the same LP cannot co-exist in both sets.

Next, we conduct several additional experiments to examine the impact of the generated images on ALPR application. Therefore, we prepare multiple sets of training data and a challenging test set.

4.2 | Evaluation metrics

To compare the performance of GAN based generators, one needs to use appropriate metrics that reveal the quality of the generated images. The Inception Score (IS) [46] is a commonly used metric, which depends on the classification scores from the pre-trained Inception v3 image classification model [47]. Higher IS value indicates better quality of the generated images. We adopt IS by replacing the Inception model with the CNN model that classifies the characters in the LP images. Unfortunately, the IS metric is inadequate to quantify the controllability and diversity in the generated images, which are very important and main contributions of this research. Therefore, it requires additional metrics for an appropriate evaluation.

Besides image quality, we aim to evaluate the proposed and competitive approaches from other aspects: (a) ability to synthesize controlled image; (b) diversity of the synthesized LP text

17519578, 2023, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/itr2.12251 by Readcube (Labtiva Inc.), Wiley Online Library on [10/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

images, and (c) usefulness/applicability of the generated images. Therefore, following the existing metrics proposed in [4, 14, 15], we propose several metrics suitable for text recognition, which we define in the following sub-sections.

4.2.1 | Correct control score (CCS)

it quantifies the ability to successfully generate the controlled image by measuring the fraction of the input conditions that the generated images satisfy correctly. In the context of generating images with given LP text, we compute it as the average Hamming similarity score of input conditions and resulting labels of the generated images:

$$CCS = \frac{1}{N}\sum_{i=1}^{N} h_s(\boldsymbol{y}_i, \hat{\boldsymbol{y}}_i)), \tag{10}$$

where N is a number of input conditions vectors $\boldsymbol{y}_i$ used for the assessment, $\hat{\boldsymbol{y}}_i$ is the predicted attributes vector of the image generated with input conditions vector $\boldsymbol{y}_i$:

$$\hat{\boldsymbol{y}}_i = (\arg\max D^j(G(\boldsymbol{z}, \boldsymbol{y}_i)))_{j=1}^{K}. \tag{11}$$

The Hamming similarity score h_s calculates the share of indices at which the corresponding labels are the same (e.g. $h_s(\{\}\mathbf{AB}C123'', \{\}\mathbf{AB}D723'') = 0.6667 \times 100\% = 66.67\%$). Indeed, for text recognition, it is straightforward to compute CCS either by human (manual) annotator or by using a highly accurate character recognizer to label the synthesized images. The CCS value indicates the percentage, where 100% means full control green of the generated image and 0% means no control at all.

4.2.2 | Diversity of the generated images (DGI)

a common expectation from the image generator is its capability to provide a diverse set of images for a given set of input conditions. The DGI evaluation metric [4] provides an important measure to quantify the variations of the generated images for a fixed input condition as well as to detect the collapsed generator. For the natural images, the multi-scale structural similarity provides a measure of DGI. To adapt this metric to our problem, we use an in-house LP matching method and compute the DGI measure as:

$$DGI = \frac{1}{M}\sum_{m=1}^{M} d\left(x_m^1, x_m^2\right). \tag{12}$$

Here, M represents the number of given image pairs, (x_m^1, x_m^2) is a pair of images, and $d(\cdot)$ is a measure of distance between two images. While higher DGI values indicate better diversity, zero DGI value indicates one unique generation possible for a given input.

TABLE 1 Evaluation of the competitive approaches with CCS and DGI metrics.

Training configuration	IS	CCS	DGI
Baseline	22.88	88.36%	0.227
Baseline+ OS	20.48	86.50%	0.253
Baseline+ DL	21.70	87.58%	0.158
Baseline + OS + DL	**23.48**	**99.80%**	**0.248**
Pix2Pix [17]	18.22	84.05%	0

4.2.3 | Impact on target applications (ITA)

besides evaluating the generator, it is equally important to realize its impact on the target applications. To this aim, we exploit the MAC-GAN generated images for the ALPR application. Therefore, following the related work [13–15], we quantify ITA with two individual scores: Total Recognition Rate (TRR) and Partial Recognition Rate (PRR). While TRR quantifies the accuracy based on complete LP text matching, PRR quantifies the accuracy at individual character level. Therefore, ITA helps to analyze the performance obtained with samples generated with different strategies.

4.3 | Results

This section briefly presents the results from two different aspects: competitive evaluation of image synthesis and the usefulness of the synthesized images for a selected ML application.

4.3.1 | Image synthesis

in order to perform a competitive evaluation[2], we provide results from different competitive approaches: (a) MAC-GAN (*Baseline*): naive training of the proposed framework; (b) Baseline trained with oversampling (*OS*) and dynamic loss (*DL*) and (c) Pix2Pix [17]: technique used by [15] to transform a rendered LP image into a realistic appeared image. OS improves balance in the training data and DL improves the effectiveness of controlled image generation. Table 1 presents the results based on the IS, CCS and DGI metrics. We observe that the proposed MAC-GAN provides the best results based on all evaluation metrics when it is correctly trained with OS and DL strategies. Particularly, it achieves significantly higher CCS to generate controlled natural images, which satisfies the most important objective (controllability) of this research. Besides, the DGI values confirm its capability to generate wide variations for the same input (control). The higher IS score from MAC-GAN confirms that it ensures better image quality compared to the competitive methods. Results from Pix2Pix [17] indicate that while the CCS score is closer to the baseline, it is not

[2] Unfortunately, we are not able to provide a state-of-the-art comparison due to the unavailable implementations of original PixTextGAN [15] and CycleWGAN [14]. Ourimplementation of those methods struggle to generate satisfying results.

17519578, 2023, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/itr2.12251 by Readcube (Labtiva Inc.), Wiley Online Library on [10/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

FIGURE 4 Illustration of controllability and diversity in the MAC-GAN synthesized images

capable to generate a variety of different appearances for the same input. However, CCS from both baseline (86.36%) and Pix2Pix (84.05%) indicate that they are highly unreliable compared to the best CCS (99.80%). Therefore, MAC-GAN trained with OS and DL significantly outperforms its competitors.

Figure 3 illustrates the LP images from the real dataset, MAC-GAN, and Pix2Pix methods for different controlled inputs (text encoded into conditions). Visual comparison with the real images indicates that MAC-GAN generated LP images are often difficult to distinguish from the real ones. Besides, they are more natural than the ones from Pix2Pix. Note that, unlike MAC-GAN, Pix2Pix generates only one image for a given input. This further emphasizes the effectiveness of the proposed method for generating fully controllable and high-quality natural images with large variations.

One of the key properties of MAC-GAN is the controllability of the image attributes. To briefly analyse this, we visualize the appearance changes of the generated LP text images with respect to the condition vectors and a given random noise vector. Figure 4 illustrates the results, where the first column shows the reference LP text images, and then each column provides a different LP text synthesized image compared to the reference. To demonstrate controllability, at each $(j+1)^{th}$ column we only change the j^{th} control vector or attribute (here j^{th} character of the LP text) with respect to the reference. The images in Figure 4 show that MAC-GAN provides perfect controllability and allows to generate images with any set of attributes that follows the training distribution of attributes. Besides, each row of the MAC-GAN generated images in Figure 3 illustrates different visual appearances (different random noise vector) for the same LP text (each column), which ensures its capacity to synthesize images with diverse appearances. These observations clarify that the MAC-GAN framework allows an ML practitioner to gain full control over the attributes to generate a large number of images for the target application. In the next sub-section, we briefly present the advantages and impact of exploiting these properties for the ALPR application.

Next, we perform an analysis of MAC-GAN's performance at different resolutions for the synthesized LP text images, see Figure 2 for visual appearances. Table 2 provides the results for the IS and CCS metrics, from which it is clear that 128x256 is the optimal resolution for the LP text images. Moreover, a decrease in resolution by half causes significant degradation of the results. We believe that the optimal choice of resolution should be made based on the target application.

TABLE 2 MAC-GAN's performance for different image resolutions

	4x8	8x16	16x32	32x64	64x128	128x256
IS	0.0	0.0	2.06	10.60	10.61	23.48
CCS	0.0	0.0	0.02	0.23	0.24	99.80

4.3.2 | Impact on ALPR performance

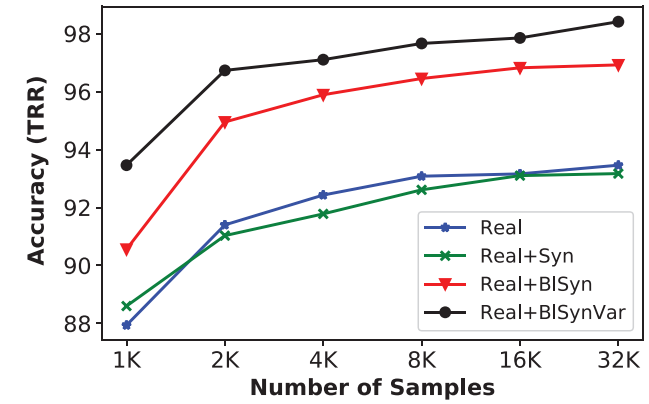
Herein, we present the scope and usefulness of the synthesized images for ALPR, which we consider as a representative of the real-world data-driven ML applications. Particularly, we aim to show the best scenario to take profit from the synthesized images, and improve the performance of the target application. Therefore, we conduct extensive experiments to train an ALPR model [11] with a large set of datasets consisting of different amounts and combinations of the real and synthesized data.

We use the 30K unique LPs (Section 4.1) and create multiple inclusive subsets of size 1K, 2K, 4K, 8K, 16K, and 32K[3] LPs. For the evaluation, we manually select 1K unique LP that does not already exist in the training sets, and that presents a challenge (under-represented characters, extreme light conditions etc.). We follow the same training procedure for any given dataset, where the training begins with a learning rate of 0.001 and gradually decreases for 20 epochs. During training, we apply (with probability 0.66) on-the-fly random data augmentation [48, 49] to modify the pixel and geometric appearance. To ensure a fair evaluation, each training begins with the same initial parameters of the ALPR model. With these subsets of LP texts, we consider different dataset constructions as follows:

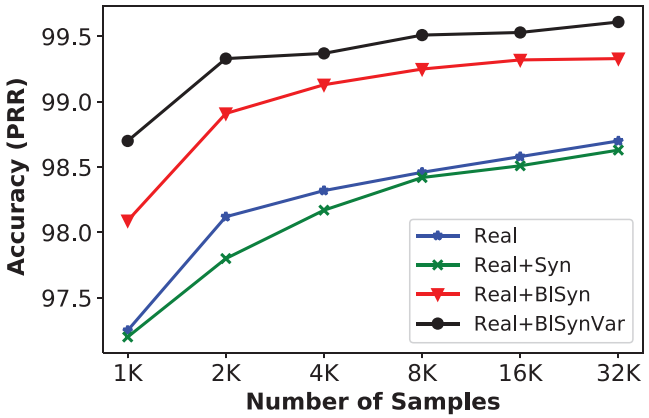
- *Real*: considers the images from the original training dataset (Section 4.1).
- *Real+Syn*: for each subset, it replaces 50% of the randomly selected real images with synthesized images of the same LP text. This helps to understand whether synthesized images can act as an exact replacement for real data.
- *Real+BlSyn*: it aims to incorporate content/class balance into the dataset. We replace 50% of the real LPs with a particular set of LPs that improves class balance within the training dataset. This helps to understand whether synthesized images can be helpful if we use them to improve the class balance.
- *Real+BlSynVar*: it aims to increase the volume of the training data by adding variations of the samples. Therefore, we use MAC-GAN to generate 5 additional images for each

[3] Originally we have 30K unique plates. Therefore, to create a 32K LP database, we randomly choose 2K additional samples from 30K LPs.

17519578, 2023, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/itr2.12251 by Readcube (Labtiva Inc.), Wiley Online Library on [10/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) **TRR**



(b) **PRR**

FIGURE 5 Illustration of the impact of using MAC-GAN synthesized images based on the ITA evaluation: (a) Total Recognition Rate (TRR) and (b) Partial Recognition Rate (PRR)

LP text of the **Real+BlSyn**. However, increasing variations enlarges the volume of the training dataset and hence applying the same (like the previous sets) training strategy may make the comparison unfair. Therefore, we maintain fairness by adjusting the number of epochs.

Figure 5 presents the results obtained from the training on 24 different datasets based on TRR and PRR. From these results, our observations are:

1. ALPR performance always increases with the amount of training data, regardless of the dataset construction procedures. Therefore, the use of image generator can be very helpful for the ML applications. Results from TRR and PRR evaluation show a similar trend.
2. Comparison among **Real** and **Real+Syn** shows that though the use of real data is preferable when possible, overall, replacing half of the real data by synthetic data with *the same attributes distribution* leads to similar, albeit slightly worse performance. Furthermore, by comparing **Real+Syn** to the points of **Real** curve with a shift of one step back, we observe improved performance when using additional synthesized data. These observations corroborate the intuition that (a) real data still has an advantage over synthetic data and (b) if no (additional) real data is available, synthetic data will be beneficial.
3. Results from **Real+BlSyn** demonstrate substantial improvement over **Real** and **Real+Syn**. **Real+BlSyn** ensures that the distribution of data attributes does not repeat the distribution of real data, but complements it to obtain a dataset with more balanced attributes. Indeed, without the attributes control, one could expect that attributes of generated data will, in the best case, repeat the distribution of attributes in real data, or even simply reflect only the most represented attributes values. In such a scenario, we cannot expect better performance than previously seen **Real+Syn** that can only approximate real data presence. The control over attributes unlocks access to the next level of performance in the dependent multi-attribute, multi-class tasks, which often suffer from imbalanced datasets. The latter observation relates equally well to the test sets, since fair performance assessment requires an adequate attributes representation.
4. Results from **Real+BlSynVar** show the superiority of MAC-GAN overcompetitive methods due to its ability to generate multiple samples with variations from the same input conditions. The large performance gap (≈5% on average compared to *Real*) highlights the importance of including variations within the training dataset. Moreover, the comparison with *Real+BlSyn*[4] confirms that although *re-balancing* the training dataset with generated data helps, variations in the synthesized data will further enhance the performance, even when using data augmentation.

The above observations confirm that synthesized images can significantly (achieved ≈5% on average) boost the performance of ML applications if the synthesizer can generate output with large control and sufficient variations. From this aspect, the proposed MAC-GAN framework shows enormous potential to be adopted for various ML applications.

5 | CONCLUSIONS

Generating high-fidelity images with controllable attributes can provide great benefits greento data-driven ML methods, particularly when data acquisition and data storing appear very challenging. To overcome this, this research proposes a novel method, called MAC-GAN, that generates high-quality images with controllable attributes and variability. While MAC-GAN leverages several existing frameworks (e.g. AC-GAN and PG-GAN) to ensure high-quality images, it introduces multi-conditioning and dynamic loss for greater control and diversity. Comprehensive experiments with different existing and newly proposed evaluation metrics confirm that MAC-GAN generated images are superior to the existing methods. Moreover, the high controllability and diversity of the synthesized images clearly demonstrate the novel dimension that MAC-GAN

[4] We assume that the existing LP image generation approaches [13–15] can achieve only up to this performance as they are unable to provide variations in the outputs for the same input.

17519578, 2023, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.12251 by Readcube (Labtiva Inc.), Wiley Online Library on [10/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

pioneers for image synthesis with GANs. Different experiments show that it helps achieve ≈5% ALPR improvement and thus demonstrate the significance of controllable image generation for a real-life data-driven ML application. In future work, one can explore the extension of this framework to a wider range of applications and to other topics, use cases, and domains. Moreover, its adaptability with different other GAN architectures should be examined. We believe that MAC-GAN can take it to the next level if it is successfully adapted for the generation of images based on very few examples per class and attribute.

ACKNOWLEDGEMENTS

This work was funded by vinci autoroutes and cyclope.ai. The authors would like to thank Mr. Antoine Meicler that was one of the main investigators of this job.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Amir Nakib https://orcid.org/0000-0001-9620-9324

REFERENCES

1. Goodfellow, I., Pouget Abadie, J., Mirza, M., Xu, B., Warde Farley, D., Ozair, S., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680. MIT Press, Cambridge (2014)
2. Brock, A., Donahue, J., Simonyan, K.: Large scale GAN training for high fidelity natural image synthesis. In: International Conference on Learning Representations. ICLR, Vienna (2019)
3. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of StyleGAN. arXiv preprint, arXiv:191204958 (2019)
4. Odena, A., Olah, C., Shlens, J.: Conditional image synthesis with auxiliary classifier gans. In: Proc. of the 34th International Conference on Machine Learning-Volume 70, pp. 2642–2651. JMLR. org, (2017)
5. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. In: International Conference on Learning Representations (ICLR). Morgan Kaufmann, San Francisco (2018)
6. Deng, L.: The MNIST database of handwritten digit images for machine learning research. IEEE Signal Process Mag. 29(6), 141–142 (2012)
7. Hinton, G.E.: Learning multiple layers of representation. Trends Cogn. Sci. 11(10), 428–434 (2007)
8. Liu, Z., Luo, P., Wang, X., Tang, X.: Large-scale celebfaces attributes (celeba) dataset. Retrieved 15 August 2018
9. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al.: Microsoft coco: Common objects in context. In: European Conference on Computer Vision, pp. 740–755. Springer, Berlin (2014)
10. Wah, C., Branson, S., Welinder, P., Perona, P., Belongie, S.: The caltech-ucsd birds-200-2011 dataset (2011)
11. Špaňhel, J., Sochor, J., Juránek, R., Herout, A., Maršík, L., Zemčík, P.: Holistic recognition of low quality license plates by cnn using track annotated data. In: IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–6. IEEE, Piscataway (2017)
12. Xu, Z., Yang, W., Meng, A., Lu, N., Huang, H., Ying, C., et al.: Towards end-to-end license plate detection and recognition: A large dataset and baseline. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 255–271. Springer, Berlin (2018)
13. Wu, C., Xu, S., Song, G., Zhang, S.: How many labeled license plates are needed? In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV), pp. 334–346. Springer, Cham (2018)
14. Wang, X., Man, Z., You, M., Shen, C.: Adversarial generation of training examples: applications to moving vehicle license plate recognition. arXiv preprint, arXiv:170703124 (2017)
15. Wu, S., Zhai, W., Cao, Y.: Pixtextgan: structure aware text image synthesis for license plate recognition. IET Image Process. 13(14), 2744–2752 (2019)
16. Regulation, P.: Regulation (eu) 2016/679 of the european parliament and of the council. Regulation (EU) 679, 2016 (2016)
17. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134. IEEE, Piscataway (2017)
18. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232. IEEE, Piscataway (2017)
19. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:14111784 (2014)
20. Nguyen, A., Clune, J., Bengio, Y., Dosovitskiy, A., Yosinski, J.: Plug & play generative networks: Conditional iterative generation of images in latent space. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4467–4477. IEEE, Piscataway (2017)
21. Jin, Y., Zhang, J., Li, M., Tian, Y., Zhu, H., Fang, Z.: Towards the automatic anime characters creation with generative adversarial networks. arXiv preprint arXiv:170805509 (2017)
22. Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., et al.: Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5907–5915. IEEE, Piscataway (2017)
23. Reed, S.E., Akata, Z., Mohan, S., Tenka, S., Schiele, B., Lee, H.: Learning what and where to draw. In: Advances in Neural Information Processing Systems, pp. 217–225. MIT Press, Cambridge (2016)
24. Tang, W., Li, T., Nian, F., Wang, M.: Mscgan: Multi-scale conditional generative adversarial networks for person image generation. arXiv preprint, arXiv:181008534 (2018)
25. Ramasinghe, S., Farazi, M., Khan, S.H., Barnes, N., Gould, S.: Rethinking conditional gan training: An approach using geometrically structured latent manifolds. Adv. Neural Infor. Process. Syst. 34, (2021)
26. Hamada, K., Tachibana, K., Li, T., Honda, H., Uchida, Y.: Full-body high-resolution anime generation with progressive structure-conditional generative adversarial networks. In: Proceedings of the European Conference on Computer Vision (ECCV). Springer, Berlin (2018)
27. Han, C., Murao, K., Noguchi, T., Kawata, Y., Uchiyama, F., Rundo, L., et al.: Learning more with less: conditional pggan-based data augmentation for brain metastases detection using highly-rough annotation on mr images. arXiv preprint, arXiv:190209856 (2019)
28. Jung, E., Luna, M., Park, S.H.: Conditional gan with an attention-based generator and a 3d discriminator for 3d medical image generation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 318–328. Springer, Cham (2021)
29. Togo, R., Ogawa, T., Haseyama, M.: Synthetic gastritis image generation via loss function-based conditional pggan. IEEE Access 7, 87448–87457 (2019)
30. Silvano, G., Ribeiro, V., Greati, V., Bezerra, A., Silva, I., Endo, P.T., et al.: Synthetic image generation for training deep learning-based automated license plate recognition systems on the brazilian mercosur standard. Design Autom. Embed. Syst. 25(2), 113–133 (2021)
31. Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., Lee, H.: Generative adversarial text to image synthesis. arXiv preprint, arXiv:160505396 (2016)
32. Zhang, M.L., Li, Y.K., Liu, X.Y.: Towards class-imbalance aware multi-label learning. In: Twenty-Fourth International Joint Conference on Artificial Intelligence. Morgan Kaufmann, San Francisco (2015)
33. Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., et al.: Stackgan++: Realistic image synthesis with stacked generative adversarial networks. IEEE Trans. Pattern Anal. Mach. Intell. 41(8), 1947–1962 (2018)

17519578, 2023, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.12251 by Readcube (Labtiva Inc.), Wiley Online Library on [10/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

34. Li, B., Qi, X., Lukasiewicz, T., Torr, P.: Controllable text-to-image generation. In: Advances in Neural Information Processing Systems, pp. 2063–2073. MIT Press, Cambridge (2019)
35. Park, H., Yoo, Y., Kwak, N.: Mc-gan: Multi-conditional generative adversarial network for image synthesis. arXiv preprint, arXiv:180501123 (2018)
36. Gao, L., Chen, D., Zhao, Z., Shao, J., Shen, H.T.: Lightweight dynamic conditional gan with pyramid attention for text-to-image synthesis. Pattern Recogn. 110, 107384 (2021)
37. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein gan. arXiv preprint, arXiv:170107875 (2017)
38. Aigner, S., Körner, M.: Futuregan: Anticipating the future frames of video sequences using spatio-temporal 3d convolutions in progressively growing gans. arXiv preprint, arXiv:181001325 (2018)
39. Souza, D.M., Ruiz, D.D.: Towards high-resolution face pose synthesis. In: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE, Piscataway (2018)
40. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein gans. In: Advances in Neural Information Processing Systems, pp. 5767–5777. MIT Press, Cambridge (2017)
41. Villani, C.: Optimal Transport: Old and New, vol. 338. Springer Science & Business Media, New York (2008)
42. Herrera, F., Charte, F., Rivera, A.J., Del Jesus, M.J.: Multilabel classification. In: Multilabel Classification, pp. 17–31. Springer, Berlin Heidelberg (2016)
43. Charte, F., Rivera, A., del Jesus, M.J., Herrera, F.: A first approach to deal with imbalance in multi-label datasets. In: International Conference on Hybrid Artificial Intelligence Systems, pp. 150–160. Springer, Cham (2013)
44. Charte, F., Rivera, A.J., del Jesus, M.J., Herrera, F.: Mlsmote: Approaching imbalanced multilabel learning through synthetic instance generation. Knowl.-Based Syst. 89, 385–397 (2015)
45. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al.: Tensorflow: A system for large-scale machine learning. In: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16), pp. 265–283. USENIX Association, Berkeley (2016)
46. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. In: Advances in Neural Information Processing Systems, pp. 2234–2242. MIT Press, Cambridge (2016)
47. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826. IEEE, Piscataway (2016)
48. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 702–703. IEEE, Piscataway (2020)
49. Jung, A.B., Wada, K., Crall, J., Tanaka, S., Graving, J., Reinders, C., et al.: imgaug. https://github.com/aleju/imgaug. Accessed 1 February 2020

How to cite this article: Shvai, N., Hasnat, A., Nakib, A.: Multiple auxiliary classifiers GAN for controllable image generation: Application to license plate recognition. IET Intell. Transp. Syst. 17, 243–254 (2023). https://doi.org/10.1049/itr2.12251

APPENDIX A

A.1 | CNN architecture of MAC-GAN

Table A1 presents the CNN architectures of the Generator and Critic networks.

TABLE A1 CNN architectures of the Generator and Critic networks.

Generator	Act.	Output shape	Params
Latent vector	-	512 x 1 x 1	-
Label 1	-	10 x 1 x 1	-
Label 2	-	10 x 1 x 1	-
Label 3	-	10 x 1 x 1	-
Label 4	-	10 x 1 x 1	-
Label 5	-	10 x 1 x 1	-
Label 6	-	10 x 1 x 1	-
Label 7	-	10 x 1 x 1	-
Conv 4 x 4	LReLU	512 x 4 x 4	4.8M
Conv 3 x 3	LReLU	512 x 4 x 4	2.4M
Upsample	-	512 x 8 x 8	-
Conv 3 x 3	LReLU	512 x 8 x 8	2.4M
Conv 3 x 3	LReLU	512 x 8 x 8	2.4M
Upsample	-	512 x 16 x 16	-
Conv 3 x 3	LReLU	512 x 16 x 16	2.4M
Conv 3 x 3	LReLU	512 x 16 x 16	2.4M
Upsample	-	512 x 32 x 32	0
Conv 3 x 3	LReLU	512 x 32 x 32	2.4M
Conv 3 x 3	LReLU	512 x 32 x 32	2.4M
Upsample	-	512 x 64 x 64	0
Conv 3 x 3	LReLU	256 x 64 x 64	1.2M
Conv 3 x 3	LReLU	256 x 64 x 64	590k
Upsample	-	256 x 128 x 128	0
Conv 3 x 3	LReLU	128 x 128 x 128	295k
Conv 3 x 3	LReLU	128 x 128 x 128	148k
Upsample	-	128 x 256 x 256	-
Conv 3 x 3	LReLU	64 x 256 x 256	74k
Conv 3 x 3	LReLU	64 x 256 x 256	37k
Conv 1 x 1	linear	3 x 256 x 256	195
Total trainable parameters			23.6M
Critic	**Act.**	**Output shape**	**Params**
Input image	-	3 x 256 x 256	-
Conv 1 x 1	LReLU	64 x 256 x 256	256
Conv 3 x 3	LReLU	64 x 256 x 256	37k
Conv 3 x 3	LReLU	128 x 256 x 256	74k
Downsample	-	128 x 128 x 128	-
Conv 3 x 3	LReLU	128 x 128 x 128	148k
Conv 3 x 3	LReLU	256 x 128 x 128	295k
Downsample	-	256 x 64 x 64	-
Conv 3 x 3	LReLU	256 x 64 x 64	590k
Conv 3 x 3	LReLU	512 x 64 x 64	1.2M
Downsample	-	512 x 32 x 32	-
Conv 3 x 3	LReLU	512 x 32 x 32	2.4M
Conv 3 x 3	LReLU	512 x 32 x 32	2.4M
Downsample	-	512 x 16 x 16	-

(Continues)

17519578, 2023, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/itr2.12251 by Readcube (Labtiva Inc.), Wiley Online Library on [10/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE A1 (Continued)

Critic	Act.	Output shape	Params
Conv 3 x 3	LReLU	512 x 16 x 16	2.4M
Conv 3 x 3	LReLU	512 x 16 x 16	2.4M
Downsample	-	512 x 8 x 8	-
Conv 3 x 3	LReLU	512 x 8 x 8	2.4M
Conv 3 x 3	LReLU	512 x 8 x 8	2.4M
Downsample	-	512 x 4 x 4	-
Minibatch stddev	-	513 x 4 x 4	-
Conv 3 x 3	LReLU	512 x 4 x 4	2.4M
Conv 4 x 4	LReLU	512 x 1 x 1	4.2M
Fully-connected	linear	1 x 1 x 1	36k
Fully-connected	softmax	10 x 1 x 1	-
Fully-connected	softmax	10 x 1 x 1	-
Fully-connected	softmax	10 x 1 x 1	-
Fully-connected	softmax	10 x 1 x 1	-
Fully-connected	softmax	10 x 1 x 1	-
Fully-connected	softmax	10 x 1 x 1	-
Fully-connected	softmax	10 x 1 x 1	-
Total trainable parameters			23.1M

A.2 | LP-100K dataset

In an effort to advance the research in these areas of multi-conditional generation and ALPR, we created an LP-100K dataset consisting of 100K generated LP by our trained generator. We observe that most publicly available image datasets are limited to green single-class labeling or multi-classes with binary attributes and are not suitable for multi-label learning in the context of adversarial generation. We hope that releasing this dataset will both permit the legitimate comparison of ALPR systems and will advance the exploration of multi-label controlled generation architectures.

17519578, 2023, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/itr2.12251 by Readcube (Labtiva Inc.), Wiley Online Library on [10/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License