upGrad

#RahoAmbitious

LJMU upGrad

Research Overview of Recent Trends in GANs, Transformers, BERT, and LSTM

Dr. Sahil Sharma, PhD
SME @ upGrad

upGrad

Today's Agenda

1. **Trends in Generative Adversarial Networks**
    1. **GANs in Images**
    2. **GANs in Videos**

2. **Trends in BERT, Transformers, and LSTM**
    1. **RNN Family**

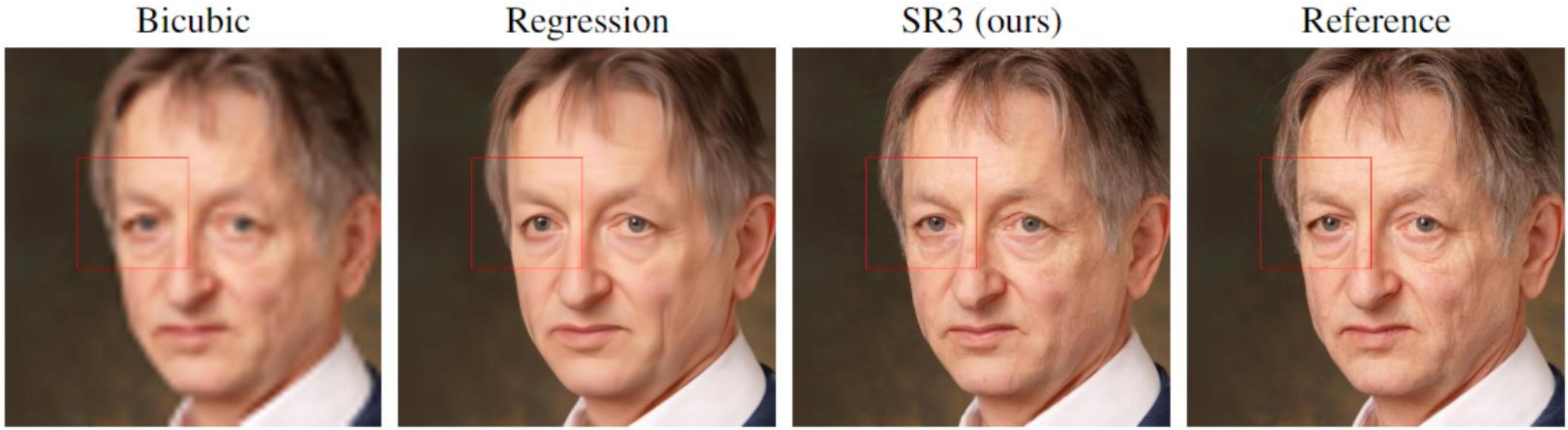
3. **Thesis Structure (Bonus)**

upGrad

GANs in Images

GANs in Image-based Research: Super Resolution

upGrad



Results of a SR3 model (64×64→512×512), trained on FFHQ

Reference: Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J. and Norouzi, M., 2021. Image super-resolution via iterative refinement. *arXiv preprint arXiv:2104.07636*.

GANs in Image-based Research: Neural Rendering

upGrad

CIPS-3D: A 3D-Aware Generator of GANs Based on Conditionally-Independent Pixel Synthesis

Peng Zhou[1], Lingxi Xie[2], Bingbing Ni[1], Qi Tian[2]
[1]Shanghai Jiao Tong University, [2]Huawei Inc.
{zhoupengcv, nibingbing}@sjtu.edu.cn, 198808xc@gmail.com,
tian.qi1@huawei.com

Figure 1. Three types of 3D-aware GANs. (a): There are apparent artifacts in the images generated by GIRAFFE [42]. (b): The images generated by pi-GAN [11] are blurred and lack details. (c): CIPS-3D can generate photo-realistic high-fidelity images. We fine-tune the base model trained on FFHQ so that the transferred model can generate other types of style images. Then we interpolate the base model and the transferred model to create a new model that can generate stylized images. CIPS-3D enables one to manipulate the pose of the stylized faces (the rightmost images) explicitly. For details, please refer to Secs. 4.4 and 4.5.

GANs in Image-based Research: : Neural Rendering

upGrad

GANcraft: Unsupervised 3D Neural Rendering of Minecraft Worlds

Zekun Hao*†, Arun Mallya*, Serge Belongie†, Ming-Yu Liu*
*NVIDIA, †Cornell University
{hz472, sjb344}@cornell.edu, {amallya, mingyul}@nvidia.com

Figure 1: Given a semantically-labeled block world as input (insets), GANCraft generates high-resolution view-consistent realistic outputs. It unsupervisedly learns to translate the input world to a realistic-looking world in the absence of paired training data across these two worlds. *Click on image to play video in web browser.*

GANs in Image-based Research

upGrad

"Our key hypothesis is that incorporating a compositional 3D scene representation into the generative model leads to more controllable image synthesis."

GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields

Michael Niemeyer[1,2] Andreas Geiger[1,2]
[1]Max Planck Institute for Intelligent Systems, Tübingen [2]University of Tübingen
{firstname.lastname}@tue.mpg.de

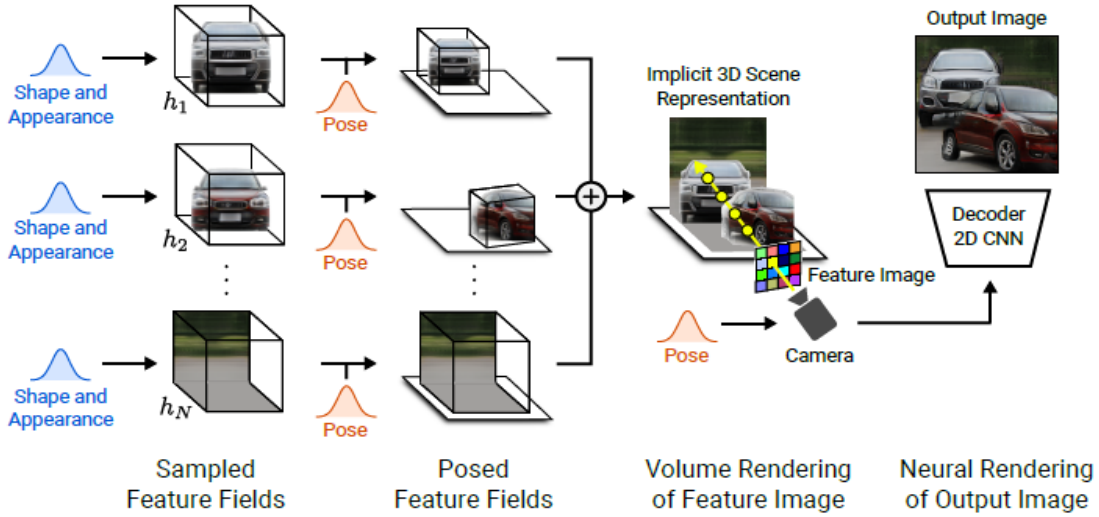


Figure 1: **Overview.** We represent scenes as compositional generative neural feature fields. For a randomly sampled camera, we volume render a feature image of the scene based on individual feature fields. A 2D neural rendering network converts the feature image into an RGB image. While training only on raw image collections, at test time we are able to control the image formation process wrt. camera pose, object poses, as well as the objects' shapes and appearances. Further, our model generalizes beyond the training data, e.g. we can synthesize scenes with more objects than were present in the training images. Note that for clarity we visualize volumes in color instead of features.

GANs in Image-based Research: Rendering

upGrad

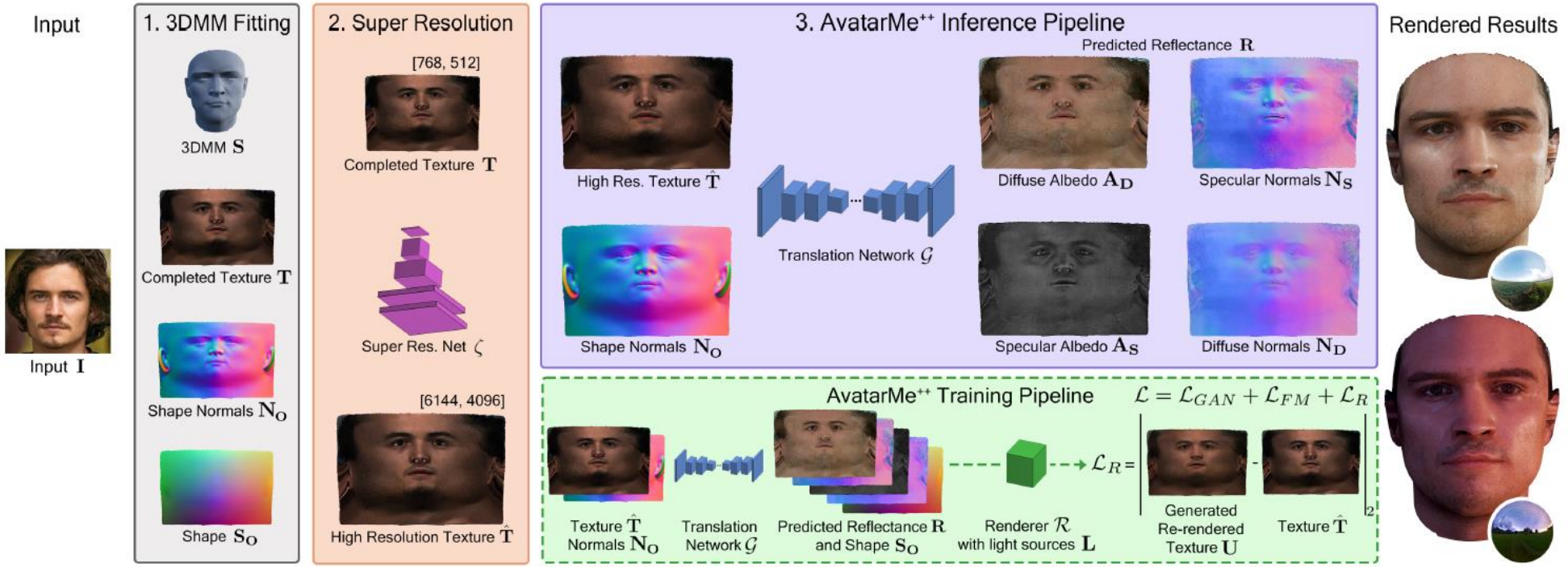


Fig. 3: Summary of the AvatarMe^{++} method: Given an "in-the-wild" image $\mathbf{I}$, we first fit a 3D Morphable Model (3DMM) to acquire the shape $\mathbf{S_O}$, texture $\mathbf{T}$ and shape normals $\mathbf{N_O}$ in UV space. Then, we upscale the texture $\mathbf{T}$ using a state-of-the-art super resolution network ζ, trained on synthetic data rendered in the texture's $\mathbf{T}$ domain. A deep network $\mathcal{G}$ is then used to transform the upscaled texture $\hat{\mathbf{T}}$ and normals $\mathbf{N_O}$) to reflectance maps, namely the diffuse albedo $\mathbf{A_D}$, specular albedo $\mathbf{A_S}$, diffuse normals $\mathbf{N_D}$ and specular normals $\mathbf{N_S}$. The deep image-translation network is trained on high-resolution captured facial BRDF, which we have made public as RealFaceDB. To train AvatarMe^{++}, we define a photorealistic differentiable rendering module $\mathcal{R}$, with subsurface-scattering and self-occlusion approximation. During training, $\mathcal{R}$ is used to create synthetic data pairs, by rendering the captured data in the target's environment $\mathbf{L}$ and random ones. The loss $\mathcal{L}$ used during training, is comprised of an adversarial loss $\mathcal{L}_{GAN}$, a feature-matching loss $\mathcal{L}_{FM}$ and our photorealistic differentiable loss $\mathcal{L}_R$. The complete high resolution (up to 6k $\times$ 4k) BRDF maps can be used for photorealistic rendering, while the specular normals $\mathbf{N_S}$ can be used to enhance the 3DMM's geometry.

AvatarMe++: Facial Shape and BRDF Inference with Photorealistic Rendering-Aware GANs, IEEE TPAMI, 2021

GANs in Image-based Research: Controllable GANs

upGrad

IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 24, 2022

AgeGAN++: Face Aging and Rejuvenation With Dual Conditional GANs

Jingkuan Song, *Senior Member, IEEE*, Jingqiu Zhang, Lianli Gao, *Member, IEEE*, Zhou Zhao, and Heng Tao Shen

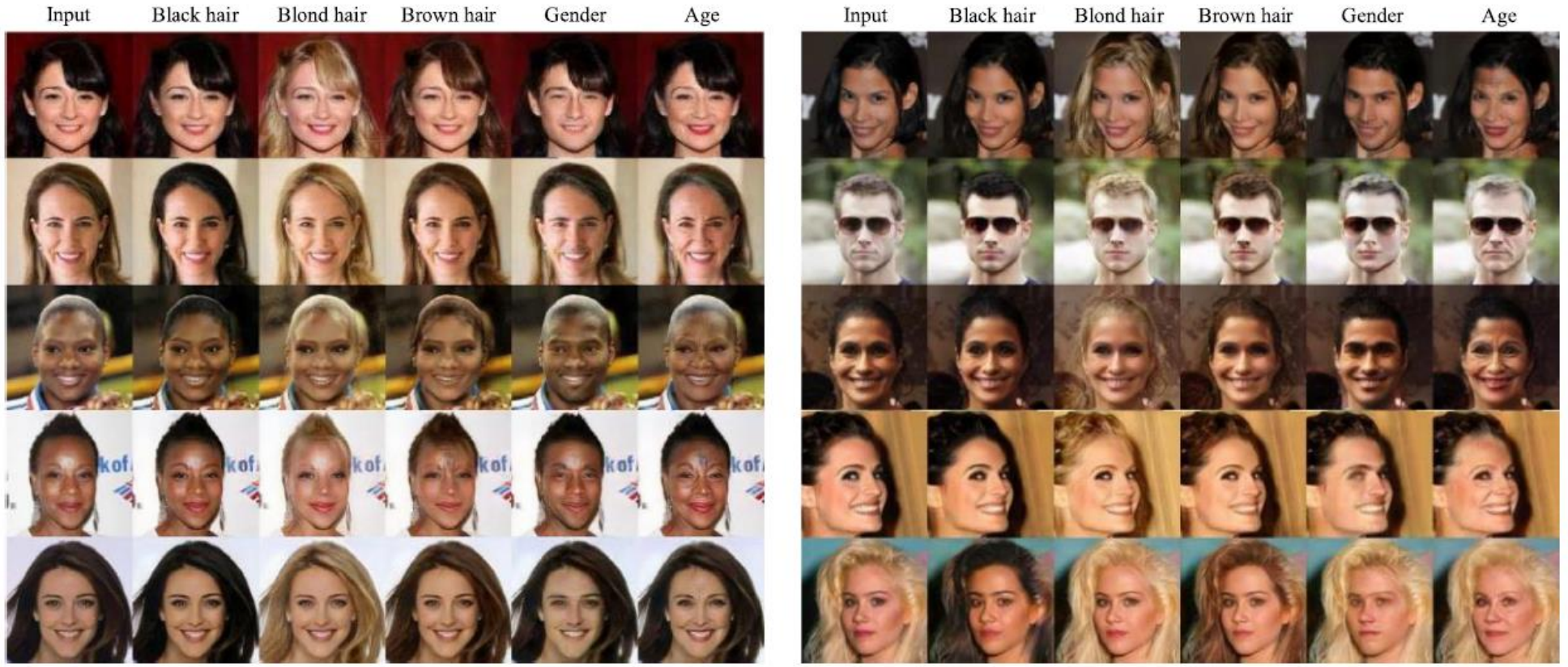


Fig. 12. Generated faces from CelebA dataset by AgeGAN++. The first column shows the input image and the last five columns are generated with the conditions of different attributes.

GANs in Image-based Research: Makeup Transfer and Removal

upGrad

Review article | Published: 10 January 2022

3D Face Reconstruction in Deep Learning Era: A Survey

Sahil Sharma & Vijay Kumar

Archives of Computational Methods in Engineering (2022) | Cite this article

Fig. 22 GAN based makeup transfer and removal [156]

upGrad

GANs in Videos

End-to-End Video-To-Speech Synthesis using Generative Adversarial Networks

Rodrigo Mira, Konstantinos Vougioukas, Pingchuan Ma, Stavros Petridis *Member, IEEE*
Björn W. Schuller *Fellow, IEEE*, Maja Pantic *Fellow, IEEE*

Video-to-speech is the process of reconstructing the audio speech from a video of a spoken utterance. Previous approaches to this task have relied on a two-step process where an intermediate representation is inferred from the video, and is then decoded into waveform audio using a vocoder or a waveform reconstruction algorithm. In this work, we propose a new end-to-end video-to-speech model based on Generative Adversarial Networks (GANs) which translates spoken video to waveform end-to-end without using any intermediate representation or separate waveform synthesis algorithm. Our model consists of an encoder-decoder architecture that receives raw video as input and generates speech, which is then fed to a waveform critic and a power critic. The use of an adversarial loss based on these two critics enables the direct synthesis of raw audio waveform and ensures its realism. In addition, the use of our three comparative losses helps establish direct correspondence between the generated audio and the input video. We show that this model is able to reconstruct speech with remarkable realism for constrained datasets such as GRID, and that it is the first end-to-end model to produce intelligible speech for LRW (Lip Reading in the Wild), featuring hundreds of speakers recorded entirely 'in the wild'. We evaluate the generated samples in two different scenarios – seen and unseen speakers – using four objective metrics which measure the quality and intelligibility of artificial speech. We demonstrate that the proposed approach outperforms all previous works in most metrics on GRID and LRW.

Neural Video Compression using GANs for Detail Synthesis and Propagation

Abstract

We present the first neural video compression method based on generative adversarial networks (GANs). Our approach significantly outperforms previous neural and non-neural video compression methods in a user study, setting a new state-of-the-art in visual quality for neural methods. We show that the GAN loss is crucial to obtain this high visual quality. Two components make the GAN loss effective: we i) synthesize detail by conditioning the generator on a latent extracted from the warped previous reconstruction to then ii) propagate this detail with high-quality flow. We find that user studies are required to compare methods, i.e., none of our quantitative metrics were able to predict all studies. We present the network design choices in detail, and ablate them with user studies.

Recent Trends in GANS:

upGrad

Diverse Generation from a Single Video Made Possible

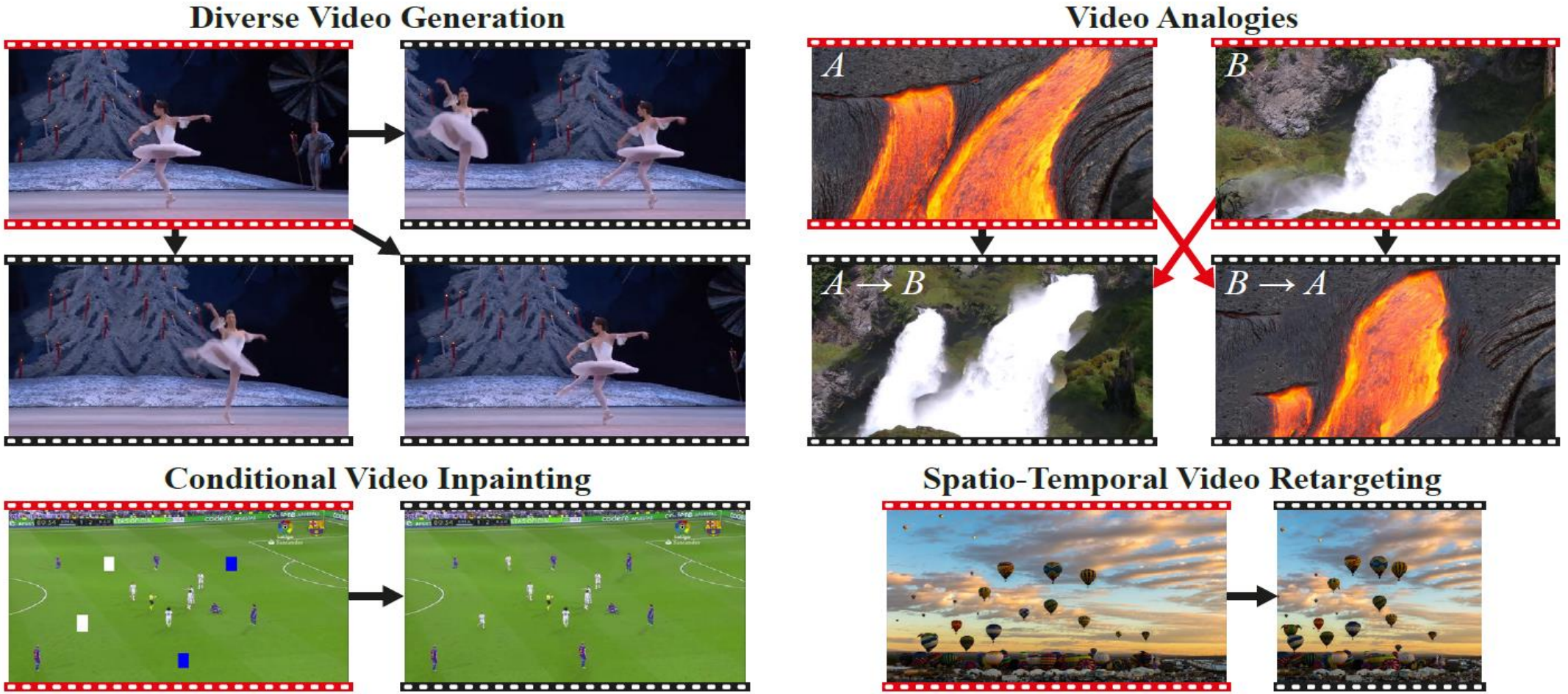


Figure 1. We adapt classical patch-based approaches as a better, much faster non-parametric alternative to single video GANs, for a variety of video generation and manipulation tasks. As we present video results, the reader is encouraged to start from the project page. Figures only present single frame examples.

Recent Trends in GANS: Unconditional Video GANs

upGrad

Video Generative Adversarial Networks: A Review

NUHA ALDAUSARI, ARCOT SOWMYA, NADINE MARCUS, and GELAREH MOHAMMADI, School of Computer Science and Engineering, University of New South Wales, Sydney, Australia

Table 1. The Reviewed Unconditional Video GANs Frameworks in Section 4.1

Publication	Conditional information	Task
VGAN [6]	no condition	Generating videos by generating background and moving objects separately.
FTGAN [56]	no condition	Generating videos from noise vectors by generating optical flow then the texture.
MoCoGAN [42]	no condition	Generating videos from noise vectors and controlling motion and content separately.
TGAN [54]	no condition	Generating videos from noise vectors by generating motion features then the texture.
TGANv2 [57]	no condition	Generating videos from noise vectors and focus on decreasing the computational costs.
DVD-GAN [48]	no condition	Generating videos from noise vectors on complex dataset (Kinetics-600).
G^3AN [58]	no condition	Generating videos from noise vectors and controlling motion and content separately.

The second column provides the type of condition while the third one gives information on the main task.

Recent Trends in GANS:

upGrad

G3AN++: Exploring Wide GANs with Complementary Feature Learning for Video Generation*

Sonam Gupta
Indian Institute of Technology Madras
Chennai, Tamil Nadu, India
cs18d005@cse.iitm.ac.in

Arti Keshari
Indian Institute of Technology Madras
Chennai, Tamil Nadu, India
cs19s008@cse.iitm.ac.in

Sukhendu Das
Indian Institute of Technology Madras
Chennai, Tamil Nadu, India
sdas@iitm.ac.in

ABSTRACT

Video generation task is a challenging problem which involves the modelling of complex real-world dynamics. Most of the existing methods have designed deep networks to tackle high-dimensional video data distributions. However, the utilization of wide networks is still under explored. Inspired by the success of wide networks in the image recognition literature, we present G3AN++, a three-stream generative adversarial network for video. The three streams are spatial, temporal and spatio-temporal processing branches. In pursuit of improving the quality of video generation, we make our network wider by splitting the spatial stream into two parallel identical branches learning complementary feature representations. We further introduce a novel adaptive masking layer to impose the complementary constraint. The masking layer encourages the parallel branches to learn distinct and richer visual features. Extensive quantitative and qualitative analysis demonstrates that our model outperforms the existing state-of-the-art methods by a significant margin on Weizmann Action, UvA-Nemo Smile and UCF101 Action datasets. Additional exploration reveals that G3AN++ is capable of disentangling the appearance and motion. We also show that the proposed method can be easily extended to solve the hard task of text-to-video generation.

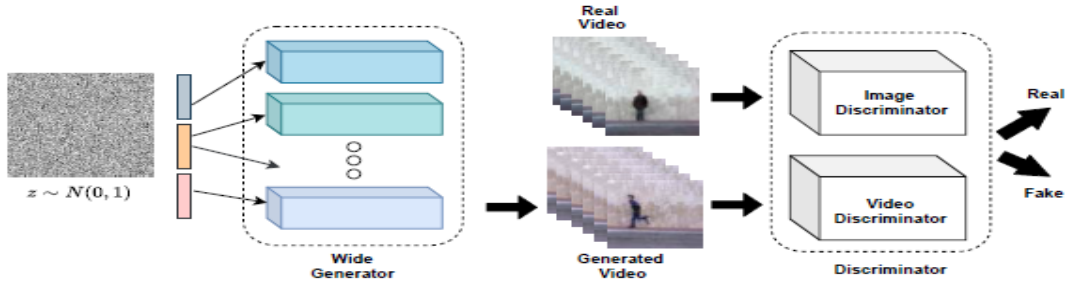


Figure 1: Diagram illustrating a high-level flowchart of the wide generative adversarial network, used in our proposed model. The generator consists of multiple parallel branches wherein each branch accepts a noise vector as input and learns distinct features.

1 INTRODUCTION

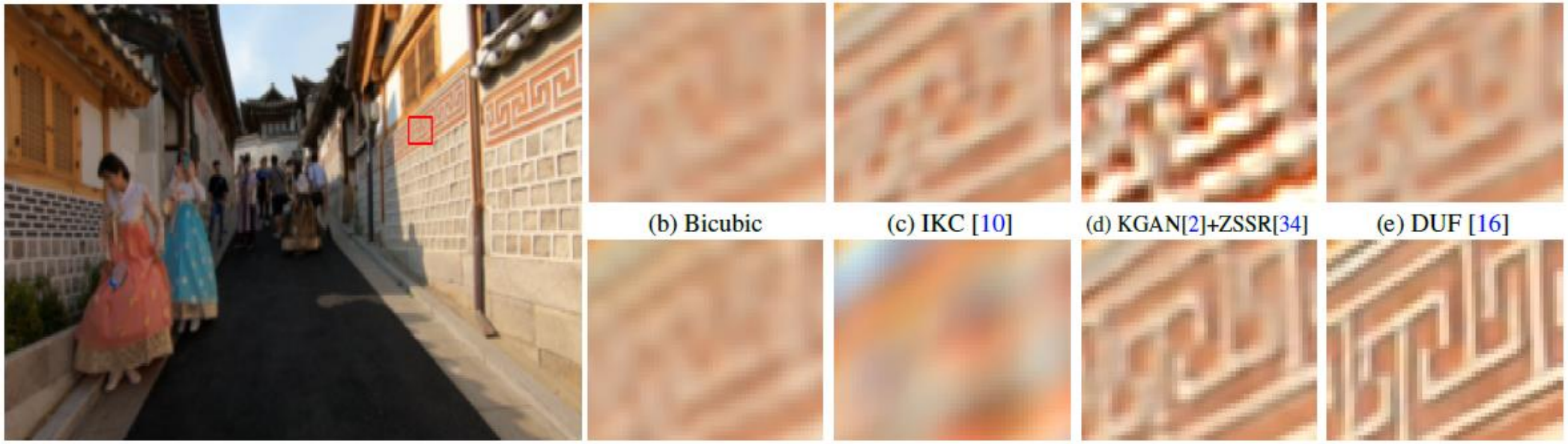
Video generation is the task of synthesizing a video from an input noise. This problem has recently received a lot of attention from the research community as it provides a means for unsupervised feature representation learning. It can also be used for data augmentation by generating new and diverse samples from the training data

Recent Trends in GANS:

upGrad

Deep Blind Video Super-resolution

Jinshan Pan[1] Haoran Bai[1] Jiangxin Dong[1] Jiawei Zhang[2] Jinhui Tang[1]*
[1]Nanjing University of Science and Technology [2]SenseTime Research



(a) Low-resolution (LR) input frame (b) Bicubic (c) IKC [10] (d) KGAN[2]+ZSSR[34] (e) DUF [16]
(f) TOFlow [43] (g) Deblur[30]+EDVR[12] (h) w/o kernel modeling (i) Ours

Figure 1. Blind video super-resolution results (×4). Existing video super-resolution algorithms usually assume the blur kernel in the degradation is known or predefined and do not model the blur kernel in the restoration process. We show that blind image super-resolution methods do not handle the video super-resolution problem well (see (c)-(d)), while existing video super-resolution methods without modeling the blur kernel does not effectively capture the intrinsic characteristics of the video super-resolution problem which thus leads to over-smoothed results (see (e)-(h)). Our algorithm explicitly estimates blur kernels from low-resolution videos, which is able to generate clearer results with finer structural details.

SMAUG: Streaming Media Augmentation Using CGANs as a Defence Against Video Fingerprinting

Alexander Vaskevich,* Thilini Dahanayaka,* Guillaume Jourjon,† Suranga Seneviratne*

* University of Sydney, Australia, † CSIRO, Space & Astronomy,
Email: {first name.last name}@sydney.edu.au, {firstname.lastname}@csiro.au

***Abstract*—Traffic fingerprinting and developing defenses against it has always been an arms race between the attackers and the defenders. The rapid evolution of deep learning methods makes developing stronger traffic fingerprinting models much easier, while overhead, latency, and deployment constraints restrict the abilities of the defenses. As such, there is always the need of coming up with novel defenses against traffic fingerprinting. In this paper, we propose SMAUG, a novel CGAN-based (Conditional Generative Adversarial Network) defense to protect video streaming traffic against fingerprinting. We first assess the performance of various GANs in video streaming traffic synthesis using multiple GAN quality metrics and show that CGAN outperforms other types of GANs such as basic GANs and WGANs (Wasserstein GAN). Our proposed defense, SMAUG, uses CGANs to synthesize video traffic flows and use those synthesized flows to camouflage the original traffic that needs protection. We compare SMAUG with other state-of-the-art defenses - FPA and d*-private methods, as well as a kernel density estimation-based baseline and show that SMAUG provides better privacy with lower overhead and delay.**

the attackers and defenders, where a more powerful traffic fingerprinting method always beats the state-of-the-art defense. With the advances in deep learning, we observe more and more fingerprinting methods being proposed [3], [12]. As such, it is required to assess and revisit existing defenses and propose new ones to circumvent the threats of traffic fingerprinting.

To this end, we propose **SMAUG** - **S**treaming **M**edia **A**ugmentation **U**sing **CGAN**s, a novel defense mechanism to defend video streaming traffic against traffic fingerprinting, using Generative Adversarial Networks (GANs) [13]. While GANs have been used extensively in computer vision tasks [14], their applications in network traffic have been explored only in limited settings [15], [16]. We leverage a Conditional Generative Adversarial Network (CGAN) [17] to generate synthetic video streaming traffic flows and use them to morph original traffic into different patterns so that an attacker cannot build a traffic classifier. To the best of our

Supervised Video-to-Video Synthesis for Single Human Pose Transfer

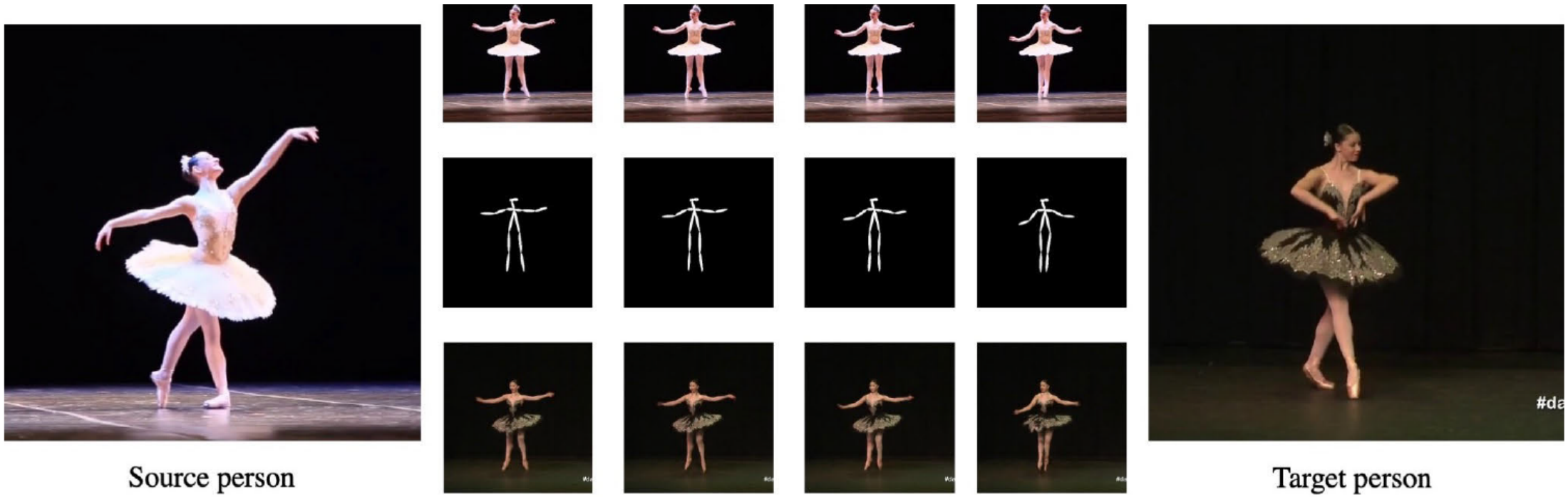


FIGURE 1. Given a dance video (the central top row) of a source person (left) and body pose frames (the central second row), our model can generate a new video (the central bottom row) of the target person (right) with the frames of source pose. The results show that our proposed method can not only produce frames with visual human appearance but also retain the details of target video, such as texture, style, color, clothes, and background.

Recent Trends in GANS: Video Summarization

upGrad

Apostolidis *et al.*: Video Summarization Using Deep Neural Networks: A Survey

ABSTRACT | Video summarization technologies aim to create a concise and complete synopsis by selecting the most informative parts of the video content. Several approaches have been developed over the last couple of decades, and the current state of the art is represented by methods that rely on modern deep neural network architectures. This work focuses on the recent advances in the area and provides a comprehensive survey of the existing deep-learning-based methods for generic video summarization. After presenting the motivation behind the development of technologies for video summarization, we formulate the video summarization task and discuss the main characteristics of a typical deep-learning-based analysis pipeline. Then, we suggest a taxonomy of the existing algorithms and provide a systematic review of the relevant literature that shows the evolution of the deep-learning-based video summarization technologies and leads to suggestions for future developments. We then report on protocols for the objective evaluation of video summarization algorithms, and we compare the performance of several deep-learning-based approaches. Based on the outcomes of these comparisons, as well as some documented considerations about the amount of annotated data and the suitability of evaluation protocols, we indicate potential future research directions.

KEYWORDS | Deep neural networks; evaluation protocols; summarization datasets; supervised learning; unsupervised learning; video summarization.

I. INTRODUCTION

In July 2015, YouTube revealed that it receives over 400 h of video content every single minute, which translates to 65.7 years' worth of content uploaded every day.[1] Since then, we are experiencing an even stronger engagement of consumers with both online video platforms and devices (e.g., smartphones and wearables) that carry powerful video recording sensors and allow instant uploading of the captured video on the Web. According to newer estimates, YouTube now receives 500 h of video per minute[2]; YouTube

Overview of the video as a set of frames

Fill Feedback Form 1

upGrad

RNN Family

- RNN, Bi-directional RNN
- LSTM, Bi-directional LSTM
- GRU, Bi-directional GRU
- Transformers
- Bidirectional Encoder Representations from Transformers (BERT)
- Generative Pretrained Transformers(GPT)

upGrad

RNN

Fake news detection: A hybrid CNN-RNN based deep learning approach

Abstract

The explosion of social media allowed individuals to spread information without cost, with little investigation and fewer filters than before. This amplified the old problem of fake news, which became a major concern nowadays due to the negative impact it brings to the communities. In order to tackle the rise and spreading of fake news, automatic detection techniques have been researched building on artificial intelligence and machine learning. The recent achievements of deep learning techniques in complex natural language processing tasks, make them a promising solution for fake news detection too. This work proposes a novel hybrid deep learning model that combines convolutional and recurrent neural networks for fake news classification. The model was successfully validated on two fake news datasets (ISO and FA-KES), achieving detection results that are significantly better than other non-hybrid baseline methods. Further experiments on the generalization of the proposed model across different datasets, had promising results.

Advancing RNN Transducer Technology for Speech Recognition

Abstract:

We investigate a set of techniques for RNN Transducers (RNN-Ts) that were instrumental in lowering the word error rate on three different tasks (Switchboard 300 hours, conversational Spanish 780 hours and conversational Italian 900 hours). The techniques pertain to architectural changes, speaker adaptation, language model fusion, model combination and general training recipe. First, we introduce a novel multiplicative integration of the encoder and prediction network vectors in the joint network (as opposed to additive). Second, we discuss the applicability of i-vector speaker adaptation to RNN-Ts in conjunction with data perturbation. Third, we explore the effectiveness of the recently proposed density ratio language model fusion for these tasks. Last but not least, we describe the other components of our training recipe and their effect on recognition performance. We report a 5.9% and 12.5% word error rate on the Switchboard and CallHome test sets of the NIST Hub5 2000 evaluation and a 12.7% WER on the Mozilla CommonVoice Italian test set.

Recent Trends in RNN Family

upGrad

Diagnosis of COVID-19 from X-rays Using Combined CNN-RNN Architecture with Transfer Learning

Abstract

The confrontation of COVID-19 pandemic has become one of the promising challenges of the world healthcare. Accurate and fast diagnosis of COVID-19 cases is essential for correct medical treatment to control this pandemic. Compared with the reverse-transcription polymerase chain reaction (RT-PCR) method, chest radiography imaging techniques are shown to be more effective to detect coronavirus. For the limitation of available medical images, transfer learning is better suited to classify patterns in medical images. This paper presents a combined architecture of convolutional neural network (CNN) and recurrent neural network (RNN) to diagnose COVID-19 from chest X-rays. The deep transfer techniques used in this experiment are VGG19, DenseNet121, InceptionV3, and Inception-ResNetV2. CNN is used to extract complex features from samples and classified them using RNN. The VGG19-RNN architecture achieved the best performance among all the networks in terms of accuracy in our experiments. Finally, Gradient-weighted Class Activation Mapping (Grad-CAM) was used to visualize class-specific regions of images that are responsible to make decision. The system achieved promising results compared to other existing systems and might be validated in the future when more samples would be available. The experiment demonstrated a good alternative method to diagnose COVID-19 for medical staff.

upGrad

LSTM

Recent Trends in RNN Family

upGrad

Bodapati, S., Bandarupally, H., Shaw, R.N. and Ghosh, A., 2021. Comparison and analysis of RNN-LSTMs and CNNs for social reviews classification. In *Advances in Applications of Data-Driven Computing* (pp. 49-59). Springer, Singapore.

Abstract

This chapter presents and compares results of simple and efficient deep learning models to perform sentiment analysis and text classification. Natural language processing is a massive domain that enables in finding solutions for many day-to-day tasks, and sentiment analysis falls under this domain. A typical sentiment analysis task can be described as a process of classifying opinions expressed in a text as positive, negative, or neutral. This chapter employs two models; one model is built using recurrent neural networks (RNNs) and long short-term memory networks (LSTMs) and the other with convolutional neural networks. In our first model, RNN-LSTMs model has been used to capture the semantic and syntactic relationships between words of a sentence with help of word2vec. In the second model, one-dimensional convolutional neural networks were used to learn structure in paragraphs of words and the techniques invariance to the specific position of features. The IMDB movie reviews dataset set is being used for sentiment analysis in both the models and the results are compared. Both the models yielded excellent results.

Recent Trends in RNN Family

upGrad

Lindemann, B., Maschler, B., Sahlab, N. and Weyrich, M., 2021. A survey on anomaly detection for technical systems using LSTM networks. *Computers in Industry*, *131*, p.103498.

Highlights

- Focusing on practical application of neural network-based detection algorithms.
- LSTM-based approaches allow dynamic and time-variant anomaly detection.
- Graph-based approaches enable unified representation of heterogeneous data.
- Transfer learning addresses frequent lack of sufficiently large and diverse datasets.
- Graph-based and transfer learning are promising, but still in early development.

Recent Trends in RNN Family

upGrad

S.I. : Higher Level Artificial Neural Network Based Intelligent Systems | Published: 20 November 2020

Air quality prediction using CT-LSTM

Jingyang Wang, Jiazheng Li, Xiaoxiao Wang, Jue Wang & Min Huang

Neural Computing and Applications **33**, 4779–4792 (2021) | Cite this article

547 Accesses | **6** Citations | Metrics

Abstract

With the development of industry, air pollution has become a serious problem. It is very important to create an air quality prediction model with high accuracy and good performance. Therefore, a new method of CT-LSTM is proposed in this paper, in which the prediction model is established by combining chi-square test (CT) and long short-term memory (LSTM) network model. CT is used to determine the influencing factors of air quality. The hourly air quality data and meteorological data from Jan. 1, 2017 to Dec. 31, 2018 are used to train the LSTM network model. The data from Jan. 1, 2019 to Dec. 31, 2019 are used to evaluate the LSTM network model. The AQI level of Shijiazhuang of Hebei Province of China from Jan. 1, 2019 to Dec. 31, 2019 is predicted with five methods (SVR, MLP, BP neural network, Simple RNN and this paper's new method). Then, a contrastive analysis of the five prediction results is made. The experimental results show that the accuracy of this new method reaches 93.7%, which is the highest in the five methods and the maximum error of this new method is 1. The correct number of days predicted by this new method is also the highest among the five methods, which is 342 days. The new method also shows good characteristics in MAE, MSE and RMSE, which makes it more accurate for people to predict the AQI level.

Recent Trends in RNN Family

upGrad

Intelligent autonomous street lighting system based on weather forecast using LSTM

Didar Tukymbekov, Ahmet Saymbetov, Madiyar Nurgaliyev, Nurzhigit Kuttybay, Gulbakhar Dosymbetova, Yeldos Svanbayev

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.energy.2021.120902

Highlights

- The architecture of an autonomous intelligent street lighting system is proposed.
- The LSTM is used to predict the energy generation of the proposed system.
- Weather forecast data are used as input for PV energy generation prediction.
- Methods for optimizing the brightness of lamps are presented.
- Probability of system failure in the simulation for a year does not exceed 0.08%.

Recent Trends in RNN Family

upGrad

Well production forecasting based on ARIMA-LSTM model considering manual operations

Dongyan Fan [a, b, c], Hai Sun [a, c], Jun Yao [a], Kai Zhang [a], Xia Yan [a], Zhixue Sun [a]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.energy.2020.119708

Highlights

- New hybrid models (ARIMA-LSTM, ARIMA-LSTM-DP) for predicting oil and gas well production time series.
- Fewer and frequent manual operations are investigated as nonlinear inputs for LSTM model.
- The new forecasting approach is applied to production time series of three actual wells.

Recent Trends in RNN Family

upGrad

Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management

H.D. Nguyen [a, b], K.P. Tran [b] ✉, S. Thomassey [b], M. Hamad [c]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.ijinfomgt.2020.102282

Highlights

- Two data-driven approached are proposed to enhance decision making better in supply chain.
- A multivariate time series forecasting is performed with a Long Short Term Memory (LSTM) network based method.
- A LSTM Autoencoder network-based method combined with a one-class support vector machine.
- The proposed approach is implemented to both benchmarking and real datasets.

Recent Trends in RNN Family

upGrad

The prediction and analysis of COVID-19 epidemic trend by combining LSTM and Markov method

Ruifang Ma, Xinqi Zheng ✉, Peipei Wang ✉, Haiyan Liu & Chunxiao Zhang

Scientific Reports **11**, Article number: 17421 (2021) | Cite this article

Abstract

Corona Virus Disease 2019 (COVID-19) has spread rapidly to countries all around the world from the end of 2019, which caused a great impact on global health and has had a huge impact on many countries. Since there is still no effective treatment, it is essential to making effective predictions for relevant departments to make responses and arrangements in advance. Under the limited data, the prediction error of LSTM model will increase over time, and its prone to big bias for medium- and long-term prediction. To overcome this problem, our study proposed a LSTM-Markov model, which uses Markov model to reduce the prediction error of LSTM model. Based on confirmed case data in the US, Britain, Brazil and Russia, we calculated the training errors of LSTM and constructed the probability transfer matrix of the Markov model by the errors. And finally, the prediction results were obtained by combining the output data of LSTM model with the prediction errors of Markov Model. The results show that: compared with the prediction results of the classical LSTM model, the average prediction error of LSTM-Markov is reduced by more than 75%, and the RMSE is reduced by more than 60%, the mean R^2 of LSTM-Markov is over 0.96. All those indicators demonstrate that the prediction accuracy of proposed LSTM-Markov model is higher than that of the LSTM model to reach more accurate prediction of COVID-19.

Recent Trends in RNN Family

upGrad

Shi, Z. and Chehade, A., 2021. A dual-LSTM framework combining change point detection and remaining useful life prediction. *Reliability Engineering & System Safety*, *205*, p.107257.

Abstract

Remaining Useful Life (RUL) prediction is a key task of Condition-based Maintenance (CBM). The massive data collected from multiple sensors enables monitoring the complex systems in near real-time. However, such multiple sensors data environments pose a challenging task of combining the sensor data to infer the quality and RUL of the system. To address this task, we propose a Dual-LSTM framework that leverages Long-Short Term Memory (LSTM) for degradation analysis and RUL prediction. The Dual-LSTM relaxes the strong assumption of the fixed change point and detects the uncertain change point unit by unit at first. Then, the Dual-LSTM predicts the health index beyond the change point which can be leveraged to calculate the RUL. The proposed Dual-LSTM (i) achieves real-time high-precision RUL prediction by connecting the change point detection and RUL prediction with the health index construction, (ii) introduces a novel one-dimension health index function, (iii) leverages historical information to achieve detection and prediction tasks by characterizing both long and short-term dependencies of sensor signals through LSTM network. The effectiveness of the proposed Dual-LSTM framework is validated and compared to state-of-art benchmark methods on two publicly available turbofan engine degradation datasets.

Recent Trends in RNN Family

upGrad

Yin, Y., Zheng, X., Hu, B., Zhang, Y. and Cui, X., 2021. EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM. *Applied Soft Computing*, *100*, p.106954.

Abstract

In recent years, graph convolutional neural networks have become research focus and inspired new ideas for emotion recognition based on EEG. Deep learning has been widely used in emotion recognition, but it is still challenging to construct models and algorithms in practical applications. In this paper, we propose a novel emotion recognition method based on a novel deep learning model (ERDL). Firstly, EEG data is calibrated by 3s baseline data and divided into segments with 6s time window, and then differential entropy is extracted from each segment to construct feature cube. Secondly, the feature cube of each segment serves as input of the novel deep learning model which fuses graph convolutional neural network (GCNN) and long-short term memories neural networks (LSTM). In the fusion model, multiple GCNNs are applied to extract graph domain features while LSTM cells are used to memorize the change of the relationship between two channels within a specific time and extract temporal features, and Dense layer is used to attain the emotion classification results. At last, we conducted extensive experiments on DEAP dataset and experimental results demonstrate that the proposed method has better classification results than the state-of-the-art methods. We attained the average classification accuracy of 90.45% and 90.60% for valence and arousal in subject-dependent experiments while 84.81% and 85.27% in subject-independent experiments.

Recent Trends in RNN Family

upGrad

Huang, F., Li, X., Yuan, C., Zhang, S., Zhang, J. and Qiao, S., 2021. Attention-emotion-enhanced convolutional LSTM for sentiment analysis. *IEEE Transactions on Neural Networks and Learning Systems*.

Abstract:

Long short-term memory (LSTM) neural networks and attention mechanism have been widely used in sentiment representation learning and detection of texts. However, most of the existing deep learning models for text sentiment analysis ignore emotion's modulation effect on sentiment feature extraction, and the attention mechanisms of these deep neural network architectures are based on word- or sentence-level abstractions. Ignoring higher level abstractions may pose a negative effect on learning text sentiment features and further degrade sentiment classification performance. To address this issue, in this article, a novel model named AEC-LSTM is proposed for text sentiment detection, which aims to improve the LSTM network by integrating emotional intelligence (EI) and attention mechanism. Specifically, an emotion-enhanced LSTM, named ELSTM, is first devised by utilizing EI to improve the feature learning ability of LSTM networks, which accomplishes its emotion modulation of learning system via the proposed emotion modulator and emotion estimator. In order to better capture various structure patterns in text sequence, ELSTM is further integrated with other operations, including convolution, pooling, and concatenation. Then, topic-level attention mechanism is proposed to adaptively adjust the weight of text hidden representation. With the introduction of EI and attention mechanism, sentiment representation and classification can be more effectively achieved by utilizing sentiment semantic information hidden in text topic and context. Experiments on real-world data sets show that our approach can improve sentiment classification performance effectively and outperform state-of-the-art deep learning-based methods significantly.

Recent Trends in RNN Family

upGrad

Srinivasu, P.N., SivaSai, J.G., Ijaz, M.F., Bhoi, A.K., Kim, W. and Kang, J.J., 2021. Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. *Sensors*, *21*(8), p.2852.

Abstract: Deep learning models are efficient in learning the features that assist in understanding complex patterns precisely. This study proposed a computerized process of classifying skin disease through deep learning based MobileNet V2 and Long Short Term Memory (LSTM). The MobileNet V2 model proved to be efficient with a better accuracy that can work on lightweight computational devices. The proposed model is efficient in maintaining stateful information for precise predictions. A grey-level co-occurrence matrix is used for assessing the progress of diseased growth. The performance has been compared against other state-of-the-art models such as Fine-Tuned Neural Networks (FTNN), Convolutional Neural Network (CNN), Very Deep Convolutional Networks for Large-Scale Image Recognition developed by Visual Geometry Group (VGG), and convolutional neural network architecture that expanded with few changes. The HAM10000 dataset is used and the proposed method has outperformed other methods with more than 85% accuracy. Its robustness in recognizing the affected region much faster with almost 2× lesser computations than the conventional MobileNet model results in minimal computational efforts. Furthermore, a mobile application is designed for instant and proper action. It helps the patient and dermatologists identify the type of disease from the affected region's image at the initial stage of the skin disease. These findings suggest that the proposed system can help general practitioners efficiently and effectively diagnose skin conditions, thereby reducing further complications and morbidity.

Recent Trends in RNN Family

upGrad

ABSTRACT

Formulae display: MathJax ?

This paper investigates the expected results of the current COVID-19 outbreak to arrivals of Chinese tourists to the USA and Australia. The growing market share of Chinese tourism and the fact that the county was the first to experience the pandemic make China a suitable proxy for predictions on global tourism. We employ data from the 2003 SARS outbreak to train a deep learning artificial neural network named Long Short Term Memory (LSTM). The neural network is calibrated for the particulars of the current pandemic. Our findings, which are cross-validated using backtesting, suggest that recovery of arrivals to pre-crisis levels can take from 6 to 12 months and this can have significant adverse effects not only on the tourism industry but also on other sectors that interact with it.

Polyzos, S., Samitas, A. and Spyridou, A.E., 2021. Tourism demand and the COVID-19 pandemic: An LSTM approach. *Tourism Recreation Research*, *46*(2), pp.175-187.

Recent Trends in RNN Family

upGrad

Priyadarshini, I. and Cotton, C., 2021. A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis. *The Journal of Supercomputing*, *77*(12), pp.13911-13932.

As the number of users getting acquainted with the Internet is escalating rapidly, there is more user-generated content on the web. Comprehending hidden opinions, sentiments, and emotions in emails, tweets, reviews, and comments is a challenge and equally crucial for social media monitoring, brand monitoring, customer services, and market research. Sentiment analysis determines the emotional tone behind a series of words may essentially be used to understand the attitude, opinions, and emotions of users. We propose a novel long short-term memory (LSTM)–convolutional neural networks (CNN)–grid search-based deep neural network model for sentiment analysis. The study considers baseline algorithms like convolutional neural networks, *K*-nearest neighbor, LSTM, neural networks, LSTM–CNN, and CNN–LSTM which have been evaluated using accuracy, precision, sensitivity, specificity, and F-1 score, on multiple datasets. Our results show that the proposed model based on hyperparameter optimization outperforms other baseline models with an overall accuracy greater than 96%.

Recent Trends in RNN Family

upGrad

Jalali, S.M.J., Ahmadian, S., Kavousi-Fard, A., Khosravi, A. and Nahavandi, S., 2021. Automated deep cnn-lstm architecture design for solar irradiance forecasting. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *52*(1), pp.54-65.

Abstract:

Accurate prediction of solar energy is an important issue for photovoltaic power plants to enable early participation in energy auction industries and cost-effective resource planning. This article introduces a new deep learning-based multistep ahead approach to improve the forecasting performance of global horizontal irradiance (GHI). A deep convolutional long short-term memory is used to extract optimal features for accurate prediction of the GHI. The performance of such deep neural networks directly depends on their architectures. To deal with this problem, a swarm evolutionary optimization method, called the sine-cosine algorithm, is applied and advanced to automatically optimize the network architecture. A three-phase modification model is proposed to increase the diversity of population and avoid premature convergence in the optimization mechanism. The performance of the proposed method is investigated using three datasets collected from three solar stations in the east of the United States. The experimental results demonstrate the superiority of the proposed method in comparison to other forecasting models.

upGrad

GRU

Recent Trends in RNN Family

upGrad

Abdelgwad, M.M., Soliman, T.H.A., Taloba, A.I. and Farghaly, M.F., 2021. Arabic aspect based sentiment analysis using bidirectional GRU based models. *Journal of King Saud University-Computer and Information Sciences*.

Abstract

Aspect-based Sentiment analysis (ABSA) accomplishes a fine-grained analysis that defines the aspects of a given document or sentence and the sentiments conveyed regarding each aspect. This level of analysis is the most detailed version that is capable of exploring the nuanced viewpoints of the reviews. The bulk of study in ABSA focuses on English with very little work available in Arabic. Most previous work in Arabic has been based on regular methods of machine learning that mainly depends on a group of rare resources and tools for analyzing and processing Arabic content such as lexicons, but the lack of those resources presents another challenge. In order to address these challenges, Deep Learning (DL)-based methods are proposed using two models based on Gated Recurrent Units (GRU) neural networks for ABSA. The first is a DL model that takes advantage of word and character representations by combining bidirectional GRU, Convolutional Neural Network (CNN), and Conditional Random Field (CRF) making up the (BGRU-CNN-CRF) model to extract the main opinionated aspects (OTE). The second is an interactive attention network based on bidirectional GRU (IAN-BGRU) to identify sentiment polarity toward extracted aspects. We evaluated our models using the benchmarked Arabic hotel reviews dataset. The results indicate that the proposed methods are better than baseline research on both tasks having 39.7% enhancement in F1-score for opinion target extraction (T2) and 7.58% in accuracy for aspect-based sentiment polarity classification (T3). Achieving F1 score of 70.67% for T2, and accuracy of 83.98% for T3.

Recent Trends in RNN Family

upGrad

Transformer-Encoder-GRU (T-E-GRU) for Chinese Sentiment Analysis on Chinese Comment Text

Binlong Zhang, Wei Zhou

Chinese sentiment analysis (CSA) has always been one of the challenges in natural language processing due to its complexity and uncertainty. Transformer has succeeded in capturing semantic features, but it uses position encoding to capture sequence features, which has great shortcomings compared with the recurrent model. In this paper, we propose T-E-GRU for Chinese sentiment analysis, which combine transformer encoder and GRU. We conducted experiments on three Chinese comment datasets. In view of the confusion of punctuation marks in Chinese comment texts, we selectively retain some punctuation marks with sentence segmentation ability. The experimental results show that T-E-GRU outperforms classic recurrent model and recurrent model with attention.

Zhang, B. and Zhou, W., 2021. Transformer-Encoder-GRU (TE-GRU) for Chinese Sentiment Analysis on Chinese Comment Text. *arXiv preprint arXiv:2108.00400*.

Recent Trends in RNN Family

upGrad

Raza, M.R., Hussain, W. and Merigó, J.M., 2021, October. Cloud Sentiment Accuracy Comparison using RNN, LSTM and GRU. In *2021 Innovations in Intelligent Systems and Applications Conference (ASYU)* (pp. 1-5). IEEE.

Abstract:

Cloud computing has become a de facto choice of many individuals and enterprises for computing solutions. In the last few years, many cloud providers appear in the market that offers the same services. It is a trivial job to choose an optimal service best suited for organisations in such a massive arms race of service providers. Existing consumer experience could help significantly build a holistic perception of their experiences that ultimately influence service adoption decisions. Sentiment analysis is an effective tool to understand consumer experience about the product or service. The sophisticated sentiment analysis could help businesses to gain a better insight and respond proactively to consumer issues. There are various methods for sentiment analysis that produces ideal results under different conditions. Therefore, it is very important to choose the right method to predict consumer's sentiment for a greatest result. In this paper we analyse the sentiment prediction accuracy of widely used neural network methods - recurrent neural network (RNN), long short-term memory (LSTM) and gated recurrent network (GRU). We use software as a service (SaaS) dataset having 6258 reviews. From analysis results we find that GRU outperforms the LSTM and RNN methods.

upGrad

Transformers

https://huggingface.co/docs/transformers/v4.15.0/en/index

upGrad

Transformers

State-of-the-art Machine Learning for Jax, Pytorch and TensorFlow

Transformers (formerly known as *pytorch-transformers* and *pytorch-pretrained-bert*) provides thousands of pretrained models to perform tasks on different modalities such as text, vision, and audio.

These models can applied on:

- Text, for tasks like text classification, information extraction, question answering, summarization, translation, text generation, in over 100 languages.

Recent Trends in RNN Family

upGrad

A Survey of Transformers

Tianyang Lin, Yuxin Wang, Xiangyang Liu, Xipeng Qiu

Transformers have achieved great success in many artificial intelligence fields, such as natural language processing, computer vision, and audio processing. Therefore, it is natural to attract lots of interest from academic and industry researchers. Up to the present, a great variety of Transformer variants (a.k.a. X-formers) have been proposed, however, a systematic and comprehensive literature review on these Transformer variants is still missing. In this survey, we provide a comprehensive review of various X-formers. We first briefly introduce the vanilla Transformer and then propose a new taxonomy of X-formers. Next, we introduce the various X-formers from three perspectives: architectural modification, pre-training, and applications. Finally, we outline some potential directions for future research.

Lin, T., Wang, Y., Liu, X. and Qiu, X., 2021. A survey of transformers. *arXiv preprint arXiv:2106.04554*.

Recent Trends in RNN Family

upGrad

Khan, S., Naseer, M., Hayat, M., Zamir, S.W., Khan, F.S. and Shah, M., 2021. Transformers in vision: A survey. *ACM Computing Surveys (CSUR)*.

Abstract

Astounding results from Transformer models on natural language tasks have intrigued the vision community to study their application to computer vision problems. Among their salient benefits, Transformers enable modeling long dependencies between input sequence elements and support parallel processing of sequence as compared to recurrent networks *e.g.*, Long short-term memory (LSTM). Different from convolutional networks, Transformers require minimal inductive biases for their design and are naturally suited as set-functions. Furthermore, the straightforward design of Transformers allows processing multiple modalities (*e.g.*, images, videos, text and speech) using similar processing blocks and demonstrates excellent scalability to very large capacity networks and huge datasets. These strengths have led to exciting progress on a number of vision tasks using Transformer networks. This survey aims to provide a comprehensive overview of the Transformer models in the computer vision discipline. We start with an introduction to fundamental concepts behind the success of Transformers i.e., self-attention, large-scale pre-training, and bidirectional feature encoding. We then cover extensive applications of transformers in vision including popular recognition tasks (*e.g.*, image classification, object detection, action recognition, and segmentation), generative modeling, multi-modal tasks (*e.g.*, visual-question answering, visual reasoning, and visual grounding), video processing (*e.g.*, activity recognition, video forecasting), low-level vision (*e.g.*, image super-resolution, image enhancement, and colorization) and 3D analysis (*e.g.*, point cloud classification and segmentation). We compare the respective advantages and limitations of popular techniques both in terms of architectural design and their experimental value. Finally, we provide an analysis on open research directions and possible future works. We hope this effort will ignite further interest in the community to solve current challenges towards the application of transformer models in computer vision.

Going Deeper With Image Transformers

Hugo Touvron, Matthieu Cord, Alexandre Sablayrolles, Gabriel Synnaeve, Hervé Jégou;
Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 32-42

Abstract

Transformers have been recently adapted for large scale image classification, achieving high scores shaking up the long supremacy of convolutional neural networks. However the optimization of vision transformers has been little studied so far. In this work, we build and optimize deeper transformer networks for image classification. In particular, we investigate the interplay of architecture and optimization of such dedicated transformers. We make two architecture changes that significantly improve the accuracy of deep transformers. This leads us to produce models whose performance does not saturate early with more depth, for instance we obtain 86.5% top-1 accuracy on Imagenet when training with no external data, we thus attain the current sate of the art with less floating-point operations and parameters. Our best model establishes the new state of the art on Imagenet with Reassessed labels and Imagenet-V2 / match frequency, in the setting with no additional training data. We share our code and models

Touvron, H., Cord, M., Sablayrolles, A., Synnaeve, G. and Jégou, H., 2021. Going deeper with image transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 32-42).

Fan, H., Xiong, B., Mangalam, K., Li, Y., Yan, Z., Malik, J. and Feichtenhofer, C., 2021. Multiscale vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 6824-6835).

Multiscale Vision Transformers

Haoqi Fan, Bo Xiong, Karttikeya Mangalam, Yanghao Li, Zhicheng Yan, Jitendra Malik, Christoph Feichtenhofer; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 6824-6835

Abstract

We present Multiscale Vision Transformers (MViT) for video and image recognition, by connecting the seminal idea of multiscale feature hierarchies with transformer models. Multiscale Transformers have several channel-resolution scale stages. Starting from the input resolution and a small channel dimension, the stages hierarchically expand the channel capacity while reducing the spatial resolution. This creates a multiscale pyramid of features with early layers operating at high spatial resolution to model simple low-level visual information, and deeper layers at spatially coarse, but complex, high-dimensional features. We evaluate this fundamental architectural prior for modeling the dense nature of visual signals for a variety of video recognition tasks where it outperforms concurrent vision transformers that rely on large scale external pre-training and are 5-10 more costly in computation and parameters. We further remove the temporal dimension and apply our model for image classification where it outperforms prior work on vision transformers. Code is available at: https://github.com/facebookresearch/SlowFast.

Ranftl, R., Bochkovskiy, A. and Koltun, V., 2021. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 12179-12188).

Vision Transformers for Dense Prediction

René Ranftl, Alexey Bochkovskiy, Vladlen Koltun; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 12179-12188

Abstract

We introduce dense prediction transformers, an architecture that leverages vision transformers in place of convolutional networks as a backbone for dense prediction tasks. We assemble tokens from various stages of the vision transformer into image-like representations at various resolutions and progressively combine them into full resolution predictions using a convolutional decoder. The transformer backbone processes representations at a constant and relatively high resolution and has a global receptive field at every stage. These properties allow the dense prediction transformer to provide finer-grained and more globally coherent predictions when compared to fully-convolutional networks. Our experiments show that this architecture yields substantial improvements on dense prediction tasks, especially when a large amount of training data is available. For monocular depth estimation, we observe an improvement of up to 28% in relative performance when compared to a state-of-the-art fully-convolutional network. When applied to semantic segmentation, dense prediction transformers set a new state of the art on ADE20K with 49.02% mIoU. We further show that the architecture can be fine-tuned on smaller datasets such as NYUv2, KITTI, and Pascal Context where it also sets the new state of the art. Our models are available at https://github.com/intel-isl/DPT.

Wu, H., Xiao, B., Codella, N., Liu, M., Dai, X., Yuan, L. and Zhang, L., 2021. Cvt: Introducing convolutions to vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 22-31).

CvT: Introducing Convolutions to Vision Transformers

Haiping Wu, Bin Xiao, Noel Codella, Mengchen Liu, Xiyang Dai, Lu Yuan, Lei Zhang; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 22-31

Abstract

We present in this paper a new architecture, named Convolutional vision Transformer (CvT), that improves Vision Transformer (ViT) in performance and efficiency by introducing convolutions into ViT to yield the best of both designs. This is accomplished through two primary modifications: a hierarchy of Transformers containing a new convolutional token embedding, and a convolutional Trasnsformer block leveraging a convolutional projection. These changes introduce desirable properties of convolutional neural networks (CNNs) to the ViT architecture (i.e. shift, scale, and distortion invariance) while maintaining the merits of Transformers (i.e. dynamic attention, global context, and better generalization). We validate CvT by conducting extensive experiments, showing that this approach achieves state-of-the-art performance over other Vision Transformers and ResNets on ImageNet-1k, with less parameters and lower FLOPs. In addition, performance gains are maintained when pretrained on larger datasets (e.g. ImageNet-22k) and fine-tuned to downstream tasks. Finally, our results show that the positional encoding, a crucial component in existing Vision Transformers, can be safely removed in our model, simplifying the design for higher resolution vision tasks. Code will be released at https://github.com/microsoft/CvT.

Recent Trends in RNN Family

upGrad

Yuan, K., Guo, S., Liu, Z., Zhou, A., Yu, F. and Wu, W., 2021. Incorporating convolution designs into visual transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 579-588).

Incorporating Convolution Designs Into Visual Transformers

Kun Yuan, Shaopeng Guo, Ziwei Liu, Aojun Zhou, Fengwei Yu, Wei Wu; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 579-588

Abstract

Motivated by the success of Transformers in natural language processing (NLP) tasks, there exist some attempts (e.g., ViT and DeiT) to apply Transformers to the vision domain. However, pure Transformer architectures often require a large amount of training data or extra supervision to obtain comparable performance with convolutional neural networks (CNNs). To overcome these limitations, we analyze the potential drawbacks when directly borrowing Transformer architectures from NLP. Then we propose a new Convolution-enhanced image Transformer (CeiT) which combines the advantages of CNNs in extracting low-level features, strengthening locality, and the advantages of Transformers in establishing long-range dependencies. Three modifications are made to the original Transformer: 1) instead of the straightforward tokenization from raw input images, we design an Image-to-Tokens (I2T) module that extracts patches from generated low-level features; 2) the feed-froward network in each encoder block is replaced with a Locally-enhanced Feed-Forward (LeFF) layer that promotes the correlation among neighboring tokens in the spatial dimension; 3) a Layer-wise Class token Attention (LCA) is attached at the top of the Transformer that utilizes the multi-level representations. Experimental results on ImageNet and seven downstream tasks show the effectiveness and generalization ability compared with previous Transformers and state-of-the-art CNNs, without requiring a large amount of training data and extra CNN teachers. Besides, CeiT models also demonstrate better convergence with 3xfewer training iterations, which can reduce the training cost significantly.

Recent Trends in RNN Family

upGrad

Raghu, M., Unterthiner, T., Kornblith, S., Zhang, C. and Dosovitskiy, A., 2021. Do vision transformers see like convolutional neural networks?. *Advances in Neural Information Processing Systems*, *34*.

Abstract

Convolutional neural networks (CNNs) have so far been the de-facto model for visual data. Recent work has shown that (Vision) Transformer models (ViT) can achieve comparable or even superior performance on image classification tasks. This raises a central question: how are Vision Transformers solving these tasks? Are they acting like convolutional networks, or learning entirely different visual representations? Analyzing the internal representation structure of ViTs and CNNs on image classification benchmarks, we find striking differences between the two architectures, such as ViT having more uniform representations across all layers. We explore how these differences arise, finding crucial roles played by self-attention, which enables early aggregation of global information, and ViT residual connections, which strongly propagate features from lower to higher layers. We study the ramifications for spatial localization, demonstrating ViTs successfully preserve input spatial information, with noticeable effects from different classification methods. Finally, we study the effect of (pretraining) dataset scale on intermediate features and transfer learning, and conclude with a discussion on connections to new architectures such as the MLP-Mixer.

LocalViT: Bringing Locality to Vision Transformers

Yawei Li, Kai Zhang, Jiezhang Cao, Radu Timofte, Luc Van Gool

We study how to introduce locality mechanisms into vision transformers. The transformer network originates from machine translation and is particularly good at modelling long-range dependencies within a long sequence. Although the global interaction between the token embeddings could be well modelled by the self-attention mechanism of transformers, what is lacking a locality mechanism for information exchange within a local region. Yet, locality is essential for images since it pertains to structures like lines, edges, shapes, and even objects.
We add locality to vision transformers by introducing depth-wise convolution into the feed-forward network. This seemingly simple solution is inspired by the comparison between feed-forward networks and inverted residual blocks. The importance of locality mechanisms is validated in two ways: 1) A wide range of design choices (activation function, layer placement, expansion ratio) are available for incorporating locality mechanisms and all proper choices can lead to a performance gain over the baseline, and 2) The same locality mechanism is successfully applied to 4 vision transformers, which shows the generalization of the locality concept. In particular, for ImageNet2012 classification, the locality-enhanced transformers outperform the baselines DeiT-T and PVT-T by 2.6\% and 3.1\% with a negligible increase in the number of parameters and computational effort. Code is available at \url{this https URL}.

Li, Y., Zhang, K., Cao, J., Timofte, R. and Van Gool, L., 2021. Localvit: Bringing locality to vision transformers. *arXiv preprint arXiv:2104.05707*.

upGrad

BERT

Recent Trends in RNN Family

upGrad

Acheampong, F.A., Nunoo-Mensah, H. and Chen, W., 2021. Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artificial Intelligence Review*, *54*(8), pp.5789-5829.

Abstract

We cannot overemphasize the essence of contextual information in most natural language processing (NLP) applications. The extraction of context yields significant improvements in many NLP tasks, including emotion recognition from texts. The paper discusses transformer-based models for NLP tasks. It highlights the pros and cons of the identified models. The models discussed include the Generative Pre-training (GPT) and its variants, Transformer-XL, Cross-lingual Language Models (XLM), and the Bidirectional Encoder Representations from Transformers (BERT). Considering BERT's strength and popularity in text-based emotion detection, the paper discusses recent works in which researchers proposed various BERT-based models. The survey presents its contributions, results, limitations, and datasets used. We have also provided future research directions to encourage research in text-based emotion detection using these models.

Recent Trends in RNN Family

upGrad

Yates, A., Nogueira, R. and Lin, J., 2021, March. Pretrained transformers for text ranking: BERT and beyond. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (pp. 1154-1156).

ABSTRACT

The goal of text ranking is to generate an ordered list of texts retrieved from a corpus in response to a query. Although the most common formulation of text ranking is search, instances of the task can also be found in many natural language processing applications. This tutorial, based on a forthcoming book, provides an overview of text ranking with neural network architectures known as transformers, of which BERT is the best-known example. The combination of transformers and self-supervised pretraining has, without exaggeration, revolutionized the fields of natural language processing (NLP), information retrieval (IR), and beyond. We provide a synthesis of existing work as a single point of entry for both researchers and practitioners. Our coverage is grouped into two categories: transformer models that perform reranking in multi-stage ranking architectures and learned dense representations that perform ranking directly. Two themes pervade our treatment: techniques for handling long documents and techniques for addressing the tradeoff between effectiveness (result quality) and efficiency (query latency). Although transformer architectures and pretraining techniques are recent innovations, many aspects of their application are well understood. Nevertheless, there remain many open research questions, and thus in addition to laying out the foundations of pretrained transformers for text ranking, we also attempt to prognosticate the future.

Recent Trends in RNN Family

upGrad

Le, N.Q.K., Ho, Q.T., Nguyen, T.T.D. and Ou, Y.Y., 2021. A transformer architecture based on BERT and 2D convolutional neural network to identify DNA enhancers from sequence information. *Briefings in bioinformatics*, *22*(5), p.bbab005.

Abstract

Recently, language representation models have drawn a lot of attention in the natural language processing field due to their remarkable results. Among them, bidirectional encoder representations from transformers (BERT) has proven to be a simple, yet powerful language model that achieved novel state-of-the-art performance. BERT adopted the concept of contextualized word embedding to capture the semantics and context of the words in which they appeared. In this study, we present a novel technique by incorporating BERT-based multilingual model in bioinformatics to represent the information of DNA sequences. We treated DNA sequences as natural sentences and then used BERT models to transform them into fixed-length numerical matrices. As a case study, we applied our method to DNA enhancer prediction, which is a well-known and challenging problem in this field. We then observed that our BERT-based features improved more than 5–10% in terms of sensitivity, specificity, accuracy and Matthews correlation coefficient compared to the current state-of-the-art features in bioinformatics. Moreover, advanced experiments show that deep learning (as represented by 2D convolutional neural networks; CNN) holds potential in learning BERT features better than other traditional machine learning techniques. In conclusion, we suggest that BERT and 2D CNNs could open a new avenue in biological modeling using sequence information.

Recent Trends in RNN Family

upGrad

Ganesh, P., Chen, Y., Lou, X., Khan, M.A., Yang, Y., Sajjad, H., Nakov, P., Chen, D. and Winslett, M., 2021. Compressing large-scale transformer-based models: A case study on bert. *Transactions of the Association for Computational Linguistics*, *9*, pp.1061-1080.

Abstract

Pre-trained Transformer-based models have achieved state-of-the-art performance for various Natural Language Processing (NLP) tasks. However, these models often have billions of parameters, and thus are too resource- hungry and computation-intensive to suit low- capability devices or applications with strict latency requirements. One potential remedy for this is model compression, which has attracted considerable research attention. Here, we summarize the research in compressing Transformers, focusing on the especially popular BERT model. In particular, we survey the state of the art in compression for BERT, we clarify the current best practices for compressing large-scale Transformer models, and we provide insights into the workings of various methods. Our categorization and analysis also shed light on promising future research directions for achieving lightweight, accurate, and generic NLP models.

Recent Trends in RNN Family

upGrad

BERT Transformer model for Detecting Arabic GPT2 Auto-Generated Tweets

Fouzi Harrag, Maria Debbah, Kareem Darwish, Ahmed Abdelali

During the last two decades, we have progressively turned to the Internet and social media to find news, entertain conversations and share opinion. Recently, OpenAI has developed a ma-chine learning system called GPT-2 for Generative Pre-trained Transformer-2, which can pro-duce deepfake texts. It can generate blocks of text based on brief writing prompts that look like they were written by humans, facilitating the spread false or auto-generated text. In line with this progress, and in order to counteract potential dangers, several methods have been pro-posed for detecting text written by these language models. In this paper, we propose a transfer learning based model that will be able to detect if an Arabic sentence is written by humans or automatically generated by bots. Our dataset is based on tweets from a previous work, which we have crawled and extended using the Twitter API. We used GPT2-Small-Arabic to generate fake Arabic Sentences. For evaluation, we compared different recurrent neural network (RNN) word embeddings based baseline models, namely: LSTM, BI-LSTM, GRU and BI-GRU, with a transformer-based model. Our new transfer-learning model has obtained an accuracy up to 98%. To the best of our knowledge, this work is the first study where ARABERT and GPT2 were combined to detect and classify the Arabic auto-generated texts.

Harrag, F., Debbah, M., Darwish, K. and Abdelali, A., 2021. Bert transformer model for detecting Arabic GPT2 auto-generated tweets. *arXiv preprint arXiv:2101.09345*.

Recent Trends in RNN Family

upGrad

Lin, J., Nogueira, R. and Yates, A., 2021. Pretrained transformers for text ranking: Bert and beyond. *Synthesis Lectures on Human Language Technologies*, *14*(4), pp.1-325.

Abstract

The goal of text ranking is to generate an ordered list of texts retrieved from a corpus in response to a query. Although the most common formulation of text ranking is search, instances of the task can also be found in many natural language processing (NLP) applications.This book provides an overview of text ranking with neural network architectures known as transformers, of which BERT (Bidirectional Encoder Representations from Transformers) is the best-known example. The combination of transformers and self-supervised pretraining has been responsible for a paradigm shift in NLP, information retrieval (IR), and beyond.

This book provides a synthesis of existing work as a single point of entry for practitioners who wish to gain a better understanding of how to apply transformers to text ranking problems and researchers who wish to pursue work in this area. It covers a wide range of modern techniques, grouped into two high-level categories: transformer models that perform reranking in multi-stage architectures and dense retrieval techniques that perform ranking directly. Two themes pervade the book: techniques for handling long documents, beyond typical sentence-by-sentence processing in NLP, and techniques for addressing the tradeoff between effectiveness (i.e., result quality) and efficiency (e.g., query latency, model and index size). Although transformer architectures and pretraining techniques are recent innovations, many aspects of how they are applied to text ranking are relatively well understood and represent mature techniques. However, there remain many open research questions, and thus in addition to laying out the foundations of pretrained transformers for text ranking, this book also attempts to prognosticate where the field is heading.

Recent Trends in RNN Family

upGrad

Exploring Transformers in Natural Language Generation: GPT, BERT, and XLNet

M. Onat Topal, Anil Bas, Imke van Heerden

Recent years have seen a proliferation of attention mechanisms and the rise of Transformers in Natural Language Generation (NLG). Previously, state-of-the-art NLG architectures such as RNN and LSTM ran into vanishing gradient problems; as sentences grew larger, distance between positions remained linear, and sequential computation hindered parallelization since sentences were processed word by word. Transformers usher in a new era. In this paper, we explore three major Transformer-based models, namely GPT, BERT, and XLNet, that carry significant implications for the field. NLG is a burgeoning area that is now bolstered with rapid developments in attention mechanisms. From poetry generation to summarization, text generation derives benefit as Transformer-based language models achieve groundbreaking results.

Topal, M.O., Bas, A. and van Heerden, I., 2021. Exploring transformers in natural language generation: Gpt, bert, and xlnet. *arXiv preprint arXiv:2102.08036*.

Point-BERT: Pre-training 3D Point Cloud Transformers with Masked Point Modeling

Xumin Yu, Lulu Tang, Yongming Rao, Tiejun Huang, Jie Zhou, Jiwen Lu

We present Point-BERT, a new paradigm for learning Transformers to generalize the concept of BERT to 3D point cloud. Inspired by BERT, we devise a Masked Point Modeling (MPM) task to pre-train point cloud Transformers. Specifically, we first divide a point cloud into several local point patches, and a point cloud Tokenizer with a discrete Variational AutoEncoder (dVAE) is designed to generate discrete point tokens containing meaningful local information. Then, we randomly mask out some patches of input point clouds and feed them into the backbone Transformers. The pre-training objective is to recover the original point tokens at the masked locations under the supervision of point tokens obtained by the Tokenizer. Extensive experiments demonstrate that the proposed BERT-style pre-training strategy significantly improves the performance of standard point cloud Transformers. Equipped with our pre-training strategy, we show that a pure Transformer architecture attains 93.8% accuracy on ModelNet40 and 83.1% accuracy on the hardest setting of ScanObjectNN, surpassing carefully designed point cloud models with much fewer hand-made designs. We also demonstrate that the representations learned by Point-BERT transfer well to new tasks and domains, where our models largely advance the state-of-the-art of few-shot point cloud classification task. The code and pre-trained models are available at this https URL

Yu, X., Tang, L., Rao, Y., Huang, T., Zhou, J. and Lu, J., 2021. Point-BERT: Pre-training 3D Point Cloud Transformers with Masked Point Modeling. *arXiv preprint arXiv:2111.14819*.

Recent Trends in RNN Family

upGrad

Rasmy, L., Xiang, Y., Xie, Z., Tao, C. and Zhi, D., 2021. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ digital medicine*, *4*(1), pp.1-13.

Abstract

Deep learning (DL)-based predictive models from electronic health records (EHRs) deliver impressive performance in many clinical tasks. Large training cohorts, however, are often required by these models to achieve high accuracy, hindering the adoption of DL-based models in scenarios with limited training data. Recently, bidirectional encoder representations from transformers (BERT) and related models have achieved tremendous successes in the natural language processing domain. The pretraining of BERT on a very large training corpus generates contextualized embeddings that can boost the performance of models trained on smaller datasets. Inspired by BERT, we propose Med-BERT, which adapts the BERT framework originally developed for the text domain to the structured EHR domain. Med-BERT is a contextualized embedding model pretrained on a structured EHR dataset of 28,490,650 patients. Fine-tuning experiments showed that Med-BERT substantially improves the prediction accuracy, boosting the area under the receiver operating characteristics curve (AUC) by 1.21–6.14% in two disease prediction tasks from two clinical databases. In particular, pretrained Med-BERT obtains promising performances on tasks with small fine-tuning training sets and can boost the AUC by more than 20% or obtain an AUC as high as a model trained on a training set ten times larger, compared with deep learning models without Med-BERT. We believe that Med-BERT will benefit disease prediction studies with small local training datasets, reduce data collection expenses, and accelerate the pace of artificial intelligence aided healthcare.

Recent Trends in RNN Family

upGrad

Ma, T., Pan, Q., Rong, H., Qian, Y., Tian, Y. and Al-Nabhan, N., 2021. T-bertsum: Topic-aware text summarization based on bert. *IEEE Transactions on Computational Social Systems*.

Abstract:

In the era of social networks, the rapid growth of data mining in information retrieval and natural language processing makes automatic text summarization necessary. Currently, pretrained word embedding and sequence to sequence models can be effectively adapted in social network summarization to extract significant information with strong encoding capability. However, how to tackle the long text dependence and utilize the latent topic mapping has become an increasingly crucial challenge for these models. In this article, we propose a topic-aware extractive and abstractive summarization model named T-BERTSum, based on Bidirectional Encoder Representations from Transformers (BERTs). This is an improvement over previous models, in which the proposed approach can simultaneously infer topics and generate summarization from social texts. First, the encoded latent topic representation, through the neural topic model (NTM), is matched with the embedded representation of BERT, to guide the generation with the topic. Second, the long-term dependencies are learned through the transformer network to jointly explore topic inference and text summarization in an end-to-end manner. Third, the long short-term memory (LSTM) network layers are stacked on the extractive model to capture sequence timing information, and the effective information is further filtered on the abstractive model through a gated network. In addition, a two-stage extractive-abstractive model is constructed to share the information. Compared with the previous work, the proposed model T-BERTSum focuses on pretrained external knowledge and topic mining to capture more accurate contextual representations. Experimental results on the CNN/Daily mail and XSum datasets demonstrate that our proposed model achieves new state-of-the-art results while generating consistent topics compared with the most advanced method.

Recent Trends in RNN Family

upGrad

FBERT: A Neural Transformer for Identifying Offensive Content

Diptanu Sarkar, Marcos Zampieri, Tharindu Ranasinghe, Alexander Ororbia

Transformer-based models such as BERT, XLNET, and XLM-R have achieved state-of-the-art performance across various NLP tasks including the identification of offensive language and hate speech, an important problem in social media. In this paper, we present fBERT, a BERT model retrained on SOLID, the largest English offensive language identification corpus available with over 1.4 million offensive instances. We evaluate fBERT's performance on identifying offensive content on multiple English datasets and we test several thresholds for selecting instances from SOLID. The fBERT model will be made freely available to the community.

Sarkar, D., Zampieri, M., Ranasinghe, T. and Ororbia, A., 2021. Fbert: A neural transformer for identifying offensive content. *arXiv preprint arXiv:2109.05074*.

Recent Trends in RNN Family

upGrad

Transformers: "The End of History" for NLP?

Anton Chernyavskiy, Dmitry Ilvovsky, Preslav Nakov

Recent advances in neural architectures, such as the Transformer, coupled with the emergence of large-scale pre-trained models such as BERT, have revolutionized the field of Natural Language Processing (NLP), pushing the state of the art for a number of NLP tasks. A rich family of variations of these models has been proposed, such as RoBERTa, ALBERT, and XLNet, but fundamentally, they all remain limited in their ability to model certain kinds of information, and they cannot cope with certain information sources, which was easy for pre-existing models. Thus, here we aim to shed light on some important theoretical limitations of pre-trained BERT-style models that are inherent in the general Transformer architecture. First, we demonstrate in practice on two general types of tasks -- segmentation and segment labeling -- and on four datasets that these limitations are indeed harmful and that addressing them, even in some very simple and naive ways, can yield sizable improvements over vanilla RoBERTa and XLNet models. Then, we offer a more general discussion on desiderata for future additions to the Transformer architecture that would increase its expressiveness, which we hope could help in the design of the next generation of deep NLP architectures.

Chernyavskiy, A., Ilvovsky, D. and Nakov, P., 2021. Transformers:" The End of History" for NLP?. *arXiv preprint arXiv:2105.00813*.

Recent Trends in RNN Family

upGrad

Knowledge Distillation from BERT Transformer to Speech Transformer for Intent Classification

Yidi Jiang, Bidisha Sharma, Maulik Madhavi, Haizhou Li

End-to-end intent classification using speech has numerous advantages compared to the conventional pipeline approach using automatic speech recognition (ASR), followed by natural language processing modules. It attempts to predict intent from speech without using an intermediate ASR module. However, such end-to-end framework suffers from the unavailability of large speech resources with higher acoustic variation in spoken language understanding. In this work, we exploit the scope of the transformer distillation method that is specifically designed for knowledge distillation from a transformer based language model to a transformer based speech model. In this regard, we leverage the reliable and widely used bidirectional encoder representations from transformers (BERT) model as a language model and transfer the knowledge to build an acoustic model for intent classification using the speech. In particular, a multilevel transformer based teacher-student model is designed, and knowledge distillation is performed across attention and hidden sub-layers of different transformer layers of the student and teacher models. We achieve an intent classification accuracy of 99.10% and 88.79% for Fluent speech corpus and ATIS database, respectively. Further, the proposed method demonstrates better performance and robustness in acoustically degraded condition compared to the baseline method.

Jiang, Y., Sharma, B., Madhavi, M. and Li, H., 2021. Knowledge Distillation from BERT Transformer to Speech Transformer for Intent Classification. *arXiv preprint arXiv:2108.02598*.

Globalizing BERT-based Transformer Architectures for Long Document Summarization

Quentin Grail, Julien Perez, Eric Gaussier

Abstract

Fine-tuning a large language model on downstream tasks has become a commonly adopted process in the Natural Language Processing (NLP) (CITATION). However, such a process, when associated with the current transformer-based (CITATION) architectures, shows several limitations when the target task requires to reason with long documents. In this work, we introduce a novel hierarchical propagation layer that spreads information between multiple transformer windows. We adopt a hierarchical approach where the input is divided in multiple blocks independently processed by the scaled dot-attentions and combined between the successive layers. We validate the effectiveness of our approach on three extractive summarization corpora of long scientific papers and news articles. We compare our approach to standard and pre-trained language-model-based summarizers and report state-of-the-art results for long document summarization and comparable results for smaller document summarization.

Grail, Q., Perez, J. and Gaussier, E., 2021, April. Globalizing BERT-based transformer architectures for long document summarization. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (pp. 1792-1810).

Non-autoregressive Transformer-based End-to-end ASR using BERT

Fu-Hao Yu, Kuan-Yu Chen

Transformer-based models have led to a significant innovation in various classic and practical subjects, including speech processing, natural language processing, and computer vision. On top of the transformer, the attention-based end-to-end automatic speech recognition (ASR) models have become a popular fashion in recent years. Specifically, the non-autoregressive modeling, which can achieve fast inference speed and comparable performance when compared to conventional autoregressive methods, is an emergent research topic. In the context of natural language processing, the bidirectional encoder representations from transformers (BERT) model has received widespread attention, partially due to its ability to infer contextualized word representations and to obtain superior performances of downstream tasks by performing only simple fine-tuning. In order to not only inherit the advantages of non-autoregressive ASR modeling, but also receive benefits from a pre-trained language model (e.g., BERT), a non-autoregressive transformer-based end-to-end ASR model based on BERT is presented in this paper. A series of experiments conducted on the AISHELL-1 dataset demonstrates competitive or superior results of the proposed model when compared to state-of-the-art ASR systems.

Yu, F.H. and Chen, K.Y., 2021. Non-autoregressive Transformer-based End-to-end ASR using BERT. *arXiv preprint arXiv:2104.04805*.

Recent Trends in RNN Family

upGrad

BEiT: BERT Pre-Training of Image Transformers

Hangbo Bao, Li Dong, Furu Wei

We introduce a self-supervised vision representation model BEiT, which stands for Bidirectional Encoder representation from Image Transformers. Following BERT developed in the natural language processing area, we propose a masked image modeling task to pretrain vision Transformers. Specifically, each image has two views in our pre-training, i.e, image patches (such as 16x16 pixels), and visual tokens (i.e., discrete tokens). We first "tokenize" the original image into visual tokens. Then we randomly mask some image patches and fed them into the backbone Transformer. The pre-training objective is to recover the original visual tokens based on the corrupted image patches. After pre-training BEiT, we directly fine-tune the model parameters on downstream tasks by appending task layers upon the pretrained encoder. Experimental results on image classification and semantic segmentation show that our model achieves competitive results with previous pre-training methods. For example, base-size BEiT achieves 83.2% top-1 accuracy on ImageNet-1K, significantly outperforming from-scratch DeiT training (81.8%) with the same setup. Moreover, large-size BEiT obtains 86.3% only using ImageNet-1K, even outperforming ViT-L with supervised pre-training on ImageNet-22K (85.2%). The code and pretrained models are available at this https URL.

Bao, H., Dong, L. and Wei, F., 2021. Beit: Bert pre-training of image transformers. *arXiv preprint arXiv:2106.08254*.

Recent Trends in RNN Family

upGrad

BERT Prescriptions to Avoid Unwanted Headaches: A Comparison of Transformer Architectures for Adverse Drug Event Detection

Beatrice Portelli, Edoardo Lenzi, Emmanuele Chersoni, Giuseppe Serra, Enrico Santus

Abstract

Pretrained transformer-based models, such as BERT and its variants, have become a common choice to obtain state-of-the-art performances in NLP tasks. In the identification of Adverse Drug Events (ADE) from social media texts, for example, BERT architectures rank first in the leaderboard. However, a systematic comparison between these models has not yet been done. In this paper, we aim at shedding light on the differences between their performance analyzing the results of 12 models, tested on two standard benchmarks. SpanBERT and PubMedBERT emerged as the best models in our evaluation: this result clearly shows that span-based pretraining gives a decisive advantage in the precise recognition of ADEs, and that in-domain language pretraining is particularly useful when the transformer model is trained just on biomedical text from scratch.

PDF

Cite

Search

Code

Portelli, B., Lenzi, E., Chersoni, E., Serra, G. and Santus, E., 2021, April. Bert prescriptions to avoid unwanted headaches: A comparison of transformer architectures for adverse drug event detection. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (pp. 1740-1747).

Recent Trends in RNN Family

upGrad

Embodied BERT: A Transformer Model for Embodied, Language-guided Visual Task Completion

Alessandro Suglia, Qiaozi Gao, Jesse Thomason, Govind Thattai, Gaurav Sukhatme

Language-guided robots performing home and office tasks must navigate in and interact with the world. Grounding language instructions against visual observations and actions to take in an environment is an open challenge. We present Embodied BERT (EmBERT), a transformer-based model which can attend to high-dimensional, multi-modal inputs across long temporal horizons for language-conditioned task completion. Additionally, we bridge the gap between successful object-centric navigation models used for non-interactive agents and the language-guided visual task completion benchmark, ALFRED, by introducing object navigation targets for EmBERT training. We achieve competitive performance on the ALFRED benchmark, and EmBERT marks the first transformer-based model to successfully handle the long-horizon, dense, multi-modal histories of ALFRED, and the first ALFRED model to utilize object-centric navigation targets.

Suglia, A., Gao, Q., Thomason, J., Thattai, G. and Sukhatme, G., 2021. Embodied bert: A transformer model for embodied, language-guided visual task completion. *arXiv preprint arXiv:2108.04927*.

Recent Trends in RNN Family

upGrad

Shah, S.M.A. and Ou, Y.Y., 2021. TRP-BERT: Discrimination of transient receptor potential (TRP) channels using contextual representations from deep bidirectional transformer based on BERT. *Computers in Biology and Medicine*, *137*, p.104821.

Abstract

Transient receptor potential (TRP) channels are non-selective cation channels that act as ion channels and are primarily found on the plasma membrane of numerous animal cells. These channels are involved in the physiology and pathophysiology of a wide variety of biological processes, including inhibition and progression of cancer, pain initiation, inflammation, regulation of pressure, thermoregulation, secretion of salivary fluid, and homeostasis of Ca^{2+} and Mg^{2+}. Increasing evidences indicate that mutations in the gene encoding TRP channels play an essential role in a broad array of diseases. Therefore, these channels are becoming popular as potential drug targets for several diseases. The diversified role of these channels demands a prediction model to classify TRP channels from other channel proteins (non-TRP channels). Therefore, we presented an approach based on the Support Vector Machine (SVM) classifier and contextualized word embeddings from Bidirectional Encoder Representations from Transformers (BERT) to represent protein sequences. BERT is a deeply bidirectional language model and a neural network approach to Natural Language Processing (NLP) that achieves outstanding performance on various NLP tasks. We apply BERT to generate contextualized representations for every single amino acid in a protein sequence. Interestingly, these representations are context-sensitive and vary for the same amino acid appearing in different positions in the sequence. Our proposed method showed 80.00% sensitivity, 96.03% specificity, 95.47% accuracy, and a 0.56 Matthews correlation coefficient (MCC) for an independent test set. We suggest that our proposed method could effectively classify TRP channels from non-TRP channels and assist biologists in identifying new potential TRP channels.

BERT meets Shapley: Extending SHAP Explanations to Transformer-based Classifiers

Enja Kokalj, Blaž Škrlj, Nada Lavrač, Senja Pollak, Marko Robnik-Šikonja

Abstract

Transformer-based neural networks offer very good classification performance across a wide range of domains, but do not provide explanations of their predictions. While several explanation methods, including SHAP, address the problem of interpreting deep learning models, they are not adapted to operate on state-of-the-art transformer-based neural networks such as BERT. Another shortcoming of these methods is that their visualization of explanations in the form of lists of most relevant words does not take into account the sequential and structurally dependent nature of text. This paper proposes the TransSHAP method that adapts SHAP to transformer models including BERT-based text classifiers. It advances SHAP visualizations by showing explanations in a sequential manner, assessed by human evaluators as competitive to state-of-the-art solutions.

Kokalj, E., Škrlj, B., Lavrač, N., Pollak, S. and Robnik-Šikonja, M., 2021, April. BERT meets shapley: Extending SHAP explanations to transformer-based classifiers. In *Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation* (pp. 16-21).

Recent Trends in RNN Family

upGrad

Charoenkwan, P., Nantasenamat, C., Hasan, M.M., Manavalan, B. and Shoombuatong, W., 2021. BERT4Bitter: a bidirectional encoder representations from transformers (BERT)-based model for improving the prediction of bitter peptides. *Bioinformatics*, *37*(17), pp.2556-2562.

Abstract

Motivation

The identification of bitter peptides through experimental approaches is an expensive and time-consuming endeavor. Due to the huge number of newly available peptide sequences in the post-genomic era, the development of automated computational models for the identification of novel bitter peptides is highly desirable.

Results

In this work, we present BERT4Bitter, a bidirectional encoder representation from transformers (BERT)-based model for predicting bitter peptides directly from their amino acid sequence without using any structural information. To the best of our knowledge, this is the first time a BERT-based model has been employed to identify bitter peptides. Compared to widely used machine learning models, BERT4Bitter achieved the best performance with an accuracy of 0.861 and 0.922 for cross-validation and independent tests, respectively. Furthermore, extensive empirical benchmarking experiments on the independent dataset demonstrated that BERT4Bitter clearly outperformed the existing method with improvements of 8.0% accuracy and 16.0% Matthews coefficient correlation, highlighting the effectiveness and robustness of BERT4Bitter. We believe that the BERT4Bitter method proposed herein will be a useful tool for rapidly screening and identifying novel bitter peptides for drug development and nutritional research.

Recent Trends in RNN Family

upGrad

BEVT: BERT Pretraining of Video Transformers

Rui Wang, Dongdong Chen, Zuxuan Wu, Yinpeng Chen, Xiyang Dai, Mengchen Liu, Yu-Gang Jiang, Luowei Zhou, Lu Yuan

This paper studies the BERT pretraining of video transformers. It is a straightforward but worth-studying extension given the recent success from BERT pretraining of image transformers. We introduce BEVT which decouples video representation learning into spatial representation learning and temporal dynamics learning. In particular, BEVT first performs masked image modeling on image data, and then conducts masked image modeling jointly with masked video modeling on video data. This design is motivated by two observations: 1) transformers learned on image datasets provide decent spatial priors that can ease the learning of video transformers, which are often times computationally-intensive if trained from scratch; 2) discriminative clues, i.e., spatial and temporal information, needed to make correct predictions vary among different videos due to large intra-class and inter-class variations. We conduct extensive experiments on three challenging video benchmarks where BEVT achieves very promising results. On Kinetics 400, for which recognition mostly relies on discriminative spatial representations, BEVT achieves comparable results to strong supervised baselines. On Something-Something-V2 and Diving 48, which contain videos relying on temporal dynamics, BEVT outperforms by clear margins all alternative baselines and achieves state-of-the-art performance with a 71.4% and 87.2% Top-1 accuracy respectively.

Wang, R., Chen, D., Wu, Z., Chen, Y., Dai, X., Liu, M., Jiang, Y.G., Zhou, L. and Yuan, L., 2021. Bevt: Bert pretraining of video transformers. *arXiv preprint arXiv:2112.01529*.

Recent Trends in RNN Family

upGrad

Syntax-BERT: Improving Pre-trained Transformers with Syntax Trees

Jiangang Bai, Yujing Wang, Yiren Chen, Yaming Yang, Jing Bai, Jing Yu, Yunhai Tong

Pre-trained language models like BERT achieve superior performances in various NLP tasks without explicit consideration of syntactic information. Meanwhile, syntactic information has been proved to be crucial for the success of NLP applications. However, how to incorporate the syntax trees effectively and efficiently into pre-trained Transformers is still unsettled. In this paper, we address this problem by proposing a novel framework named Syntax-BERT. This framework works in a plug-and-play mode and is applicable to an arbitrary pre-trained checkpoint based on Transformer architecture. Experiments on various datasets of natural language understanding verify the effectiveness of syntax trees and achieve consistent improvement over multiple pre-trained models, including BERT, RoBERTa, and T5.

Bai, J., Wang, Y., Chen, Y., Yang, Y., Bai, J., Yu, J. and Tong, Y., 2021. Syntax-BERT: Improving pre-trained transformers with syntax trees. *arXiv preprint arXiv:2103.04350*.

Recent Trends in RNN Family

upGrad

BERT got a Date: Introducing Transformers to Temporal Tagging

Satya Almasian, Dennis Aumiller, Michael Gertz

Temporal expressions in text play a significant role in language understanding and correctly identifying them is fundamental to various retrieval and natural language processing systems. Previous works have slowly shifted from rule-based to neural architectures, capable of tagging expressions with higher accuracy. However, neural models can not yet distinguish between different expression types at the same level as their rule-based counterparts. In this work, we aim to identify the most suitable transformer architecture for joint temporal tagging and type classification, as well as, investigating the effect of semi-supervised training on the performance of these systems. Based on our study of token classification variants and encoder-decoder architectures, we present a transformer encoder-decoder model using the RoBERTa language model as our best performing system. By supplementing training resources with weakly labeled data from rule-based systems, our model surpasses previous works in temporal tagging and type classification, especially on rare classes. Our code and pre-trained experiments are available at: this https URL

Almasian, S., Aumiller, D. and Gertz, M., 2021. BERT got a Date: Introducing Transformers to Temporal Tagging. *arXiv preprint arXiv:2109.14927*.

Speech Detection using RoBERTa

upGrad

From Universal Language Model to Downstream Task: Improving RoBERTa-Based Vietnamese Hate Speech Detection

Publisher: IEEE Cite This PDF

Quang Huu Pham ; Viet Anh Nguyen ; Linh Bao Doan ; Ngoc N. Tran ; Ta Minh Thanh **All Authors**

Abstract:

Natural language processing (NLP) is a fast-growing field of artificial intelligence. Since the Transformer [32] was introduced by Google in 2017, a large number of language models such as BERT, GPT, and ELMo have been inspired by this architecture. These models were trained on huge datasets and achieved state-of-the-art results on natural language understanding. However, fine-tuning a pre-trained language model on much smaller datasets for downstream tasks requires a carefully-designed pipeline to mitigate problems of the datasets such as lack of training data and imbalanced data. In this paper, we propose a pipeline to adapt the general-purpose RoBERTa language model to a specific text classification task: Vietnamese Hate Speech Detection. We first tune the PhoBERT [1] [9] on our dataset by re-training the model on the Masked Language Model (MLM) task; then, we employ its encoder for text classification. In order to preserve pre-trained weights while learning new feature representations, we further utilize different training techniques: layer freezing, block-wise learning rate, and label smoothing. Our experiments proved that our proposed pipeline boosts the performance significantly, achieving a new state-of-the-art on Vietnamese Hate Speech Detection (HSD) campaign [2] with 0.7221 F1 score.

Multilingual Machine Comprehension

upGrad

[Submitted on 2 Jun 2020]

BERT Based Multilingual Machine Comprehension in English and Hindi

Somil Gupta, Nilesh Khade

Multilingual Machine Comprehension (MMC) is a Question-Answering (QA) sub-task that involves quoting the answer for a question from a given snippet, where the question and the snippet can be in different languages. Recently released multilingual variant of BERT (m-BERT), pre-trained with 104 languages, has performed well in both zero-shot and fine-tuned settings for multilingual tasks; however, it has not been used for English-Hindi MMC yet. We, therefore, present in this article, our experiments with m-BERT for MMC in zero-shot, mono-lingual (e.g. Hindi Question-Hindi Snippet) and cross-lingual (e.g. English QuestionHindi Snippet) fine-tune setups. These model variants are evaluated on all possible multilingual settings and results are compared against the current state-of-the-art sequential QA system for these languages. Experiments show that m-BERT, with fine-tuning, improves performance on all evaluation settings across both the datasets used by the prior model, therefore establishing m-BERT based MMC as the new state-of-the-art for English and Hindi. We also publish our results on an extended version of the recently released XQuAD dataset, which we propose to use as the evaluation benchmark for future research.

Comments: Submitted for review to the Special Issue on Deep Learning of ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)

Subjects: **Computation and Language (cs.CL)**

Cite as: arXiv:2006.01432 **[cs.CL]**

(or arXiv:2006.01432v1 **[cs.CL]** for this version)

https://doi.org/10.48550/arXiv.2006.01432

Mono vs Multilingual BERT

upGrad

[Submitted on 24 Mar 2022]

Mono vs Multilingual BERT: A Case Study in Hindi and Marathi Named Entity Recognition

Onkar Litake, Maithili Sabane, Parth Patil, Aparna Ranade, Raviraj Joshi

Named entity recognition (NER) is the process of recognising and classifying important information (entities) in text. Proper nouns, such as a person's name, an organization's name, or a location's name, are examples of entities. The NER is one of the important modules in applications like human resources, customer support, search engines, content classification, and academia. In this work, we consider NER for low-resource Indian languages like Hindi and Marathi. The transformer-based models have been widely used for NER tasks. We consider different variations of BERT like base-BERT, RoBERTa, and AlBERT and benchmark them on publicly available Hindi and Marathi NER datasets. We provide an exhaustive comparison of different monolingual and multilingual transformer-based models and establish simple baselines currently missing in the literature. We show that the monolingual MahaRoBERTa model performs the best for Marathi NER whereas the multilingual XLM-RoBERTa performs the best for Hindi NER. We also perform cross-language evaluation and present mixed observations.

Comments: Accepted at ICMISC 2022

Subjects: **Computation and Language (cs.CL)**; Machine Learning (cs.LG)

Cite as: arXiv:2203.12907 **[cs.CL]**

(or arXiv:2203.12907v1 **[cs.CL]** for this version)

https://doi.org/10.48550/arXiv.2203.12907

BERT of all trades, master of some

Denis Gordeev, Olga Lykova

Abstract

This paper describes our results for TRAC 2020 competition held together with the conference LREC 2020. Our team name was Ms8qQxMbnjJMgYcw. The competition consisted of 2 subtasks in 3 languages (Bengali, English and Hindi) where the participants' task was to classify aggression in short texts from social media and decide whether it is gendered or not. We used a single BERT-based system with two outputs for all tasks simultaneously. Our model placed first in English and second in Bengali gendered text classification competition tasks with 0.87 and 0.93 in F1-score respectively.

Anthology ID: 2020.trac-1.15
Volume: Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying
Month: May
Year: 2020
Address: Marseille, France
Venues: LREC | TRAC | WS
SIG: –
Publisher: European Language Resources Association (ELRA)
Note: –
Pages: 93–98
Language: English
URL: https://aclanthology.org/2020.trac-1.15

Thesis Structure

upGrad

Fill Feedback Form 2

Any Questions?

Thank you!

Let’s connect on LinkedIn:
https://www.linkedin.com/in/sahil301290/