The Use of Machine Learning in Computational Fluid Dynamics

Literature Review

October 2020

CDL7416-2021

Carl Wilson [U0370630]

u0370630@unimail.hud.ac.uk

1. INTRODUCTION & SCOPE

1.1 Introduction

Computational Fluid Dynamics [CFD] can provide high-fidelity aerodynamic modelling of complex geometries. The drawback of these methods is solving time and computational expense. The solve time for a steady state set of boundary conditions in industry would be measured in the order of hours, with simplified models sometimes being measured in minutes. This makes iterative design slow and expensive.

Artificial Intelligence [AI] has been introduced in recent years to optimize CFD workflows through Reinforcement Learning, ensuring effective use of CFD solve time and exploration of design space to provide a robust design. In these instances, where the CFD solver is still used, this can lead to a significant computational overhead.

This paper will explore the various methods by which machine learning can be used to reduce this time through surrogate Reduced Order Models [ROMs].

1.2 Scope

The scope of this paper is to explore how Machine Learning [ML] algorithms, specifically how Deep Neural Networks [DNNs] can be used to accelerate the design optimization process by means of a surrogate ROM. The paper will be comparing different DNN architectures considering their specific application and their accuracy in providing flow field representations [typically velocity fields] for a given 2D or 3D geometry. The scope will be open to laminar and turbulent flows, it will also be open to both compressible and non-compressible fluids as well as steady and un-steady applications.

The scope will not include the use of ML algorithms for in-loop optimisation of a design by way of Reinforcement Learning, as discussed in the introduction.

The paper will be limited strictly to using DL methods involving DNNs. It will also focus in specifically on aerodynamic component designs and CFD flow field information as opposed to heat transfer analysis or multi-phase fluid flows.

The paper will be split into three discrete sections. The core literature review will discuss several papers with varying problems and solutions associated with the topic along with an evaluation of the research, providing critical analysis. This will be followed by a discussion on the Professional, Legal, Ethical and Social [PLES] issues surrounding this specific application of AI. Finally, there will be a short section on the conclusions, limitations and potential for further work.

2. LITERATURE REVIEW

The literature review will focus on four key areas: the geometric representation and boundary conditions [input], the network architecture, the network training and finally the resulting output.

There was a common goal throughout the papers reviewed: reduce analysis time, leading to quick design optimisation and turn-around. However, each paper tackled different geometries. From primitive 2D shapes and more complex automotive car shapes such as the paper by Guo et al. (2016), to aerofoils by both Bhatnager et al. (2019) and Wu et al. (2020), complex smoke and fluid dispersion Kim et al. (2019) and larger wind farm wake modelling by Zhang & Zhao (2020).

3.1 Geometric Representations [input]

As with CFD, the geometric representation is a key step in the solving process. Developing the geometry and meshing ahead of solving can significantly change the outcome of the results with CFD. Likewise, surrogate models pay close attention to the input and geometric representations. Signed Distance Function [SDF] over a cartesian co-ordinate system was used by both Bhatnager et al. (2019) and Guo et al. (2016) as this could easily be interpreted by a neural network. SDF is particularly useful in CFD type problems as a discrete boundary can be created between the component geometry and the fluid flow field. In the case of Wu et al. (2020) the Hicks-Henne method of parameterising the aerofoil to a vector was employed. This was specifically selected due its ability to control the aerofoil surface and a create a smooth profile. The Zhang & Zhao (2020) paper did not make geometric representations, instead providing a series of wind turbine control parameters. The specific drawback of this approach in not capturing the interaction between each turbine.

3.2 The Network

In all instances reviewed, Convolutional Neural Networks [CNN's] were used within the network. CNN's or *Convnets* are suited to vision-based tasks in deep learning and are typically used in image classification. This is because CNN's do not learn global representations, they focus on small segments within an image – local patterns (Chollet, 2017). This can be observed in Figure 1.

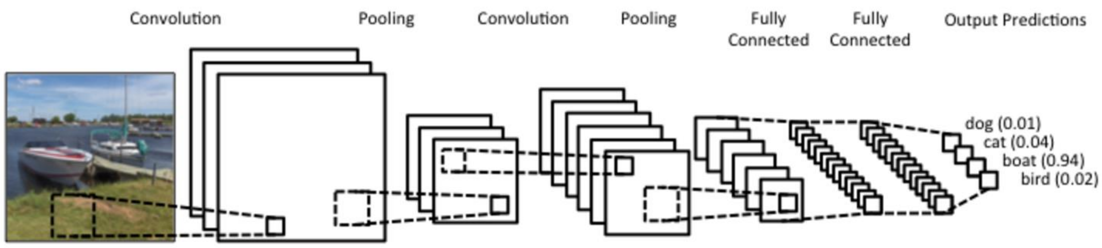

Figure 1 Typical Architecture of a CNN (Gandhi, 2018)

Encoders sometimes referred to as Variational Auto Encoders [VAE] were used by both Bhatnager et al. (2019) and Guo et al. (2016). Encoders are generative models, using CNN layers, to create an output of the same shape as the input through a decoder. This makes encoders an obvious choice for 2D based tasks such as image classification and for CFD surrogacy. One of the key differences between the two papers is the type of decoder used. A shared encoder and decoder was used by Bhatnager et al. (2019) for the x-velocity, y-velocity and pressure output vectors. However, Guo et al. (2016) used individual decoders for each of the three outputs.

A Generative Adversarial Network [GAN] was used by Wu et al. (2020) which was later referred to as the *flow field structure GAN [ffsGAN]*. GAN's are made of two key components: the generator and the discriminator. Both of which featured CNN layers for both encoding and decoding the input and output representations. The generator creates a new output based on the latent representations created by the CNN layers of the neural network, with the discriminator determining if the image is from the original data set.

As Zhang & Zaho (2020) used a set of parameters as inputs, as opposed to vectorised geometric representations, they opted for a long short-term memory [LSTM] network. This type of network is particularly strong with time-series data, which is well suited to a non-steady CFD problem such as the wind farm ROM Zhang & Zhao (2020) were attempting to emulate.

2.3 *Training*

All cases were supervised learning problems, where targets were provided and the model optimisers working to minimise the difference between the result and the target via mean squared error [MSE]. Typically a stochastic gradient decent [SGD] was used, except in the cases of Kim et al. (2019) and Zhang & Zhao (2020) where an adaptive momentum [ADAM] to ensure the global minima was obtained and the optimisation routine did not become caught in a local minima.

Guo et al. (2016) was able to create a large data set with 20,000 random variations of five primitive shapes [totalling 100,000 instances] being created to train the network and 2,000 samples for each of the five shapes [totalling 10,000 instances] being used for validation. Additionally, a series of more complex automotive car shapes were used for testing as a blind data set.

The training and validation split varied between the studies, with Bhatnager et al. (2019) using 85% of the dataset for training the network and 15% for validation using k-fold cross-validation. There were also several “unseen” aerofoil shapes for final testing of the model.

The GAN trained by Wu et al. (2020) followed a broader optimisation of the network, studying different hyper-parameters, L1 loss functions, filter sizes and numbers of CNN layers. However, this was done on 500 slightly modified aerofoil geometries, with only 16 test cases. The training and test split in this instance would likely lead to over-fitting on the training data with little test data for robust validation in understanding how well the network was generalising.

The LSTM developed by Zhang & Zhao (2020) used 710 training cases by taking a steady-state snap-shot each 1s for 180 different flow scenarios over a period of 1100s [the first 400s were not used]. The data was split: 64% training, 21% validation and 15% for blind testing. A grid-search was completed across the various hyper-parameters and drop out was used to control over-fitting.

2.4 Results

In each paper reviewed a ground truth [GT] of high-fidelity data was developed via CFD. The type of CFD varied between Reynolds Averaged Navier Stokes [RANS], Lattice Boltzmann Method [LBM] and Large Eddy Simulation [LES] with LBM and LES being employed by Guo et al. (2016) and Zhang & Zhao (2020) respectively.

In the instance where LBM was used it was found that SDF could provide accurate geometric representations in comparison to binary classification with both shared and separated encoders. A comparison was made between CNN and *patchwise* linear regression and found that CNN was more accurate to the GT. However, the error increased by a factor of five when moving from 2D primitive shapes to higher complexity automotive car profiles. With the network training time being amortised across multiple batches of solves, a significantly speed up was demonstrated.

It took around 3 hours to train the surrogate model in the case of Wu et al. (2020), this is a significant speed up as a RANS based solver can take several hours to complete one specific set of boundary conditions. With the model trained, new boundary conditions can be assessed in near real-time. The Mean Absolute Error [MAE] values indicated that the model was suitable for reasonable approximations to CFD, with values below 1%.

Kim et al. (2019) showed a MAE plot of velocity penetration for obstacles in a smoke stream. It was found that the MAE increased for interpolated examples three-fold, however remained relatively low. This gives an indication of how well the model is generalising away from the training data. Very little other hard data is provided for accuracy on training and validation, instead focussing on reductions in time and data compression. Visually there are observable differences between the surrogate model outputs and the GT.

The greatest reduction is claimed by Zhang & Zhao (2020) with multiple orders of magnitude difference in CPU hours between the ROM and CFD, it should be noted however that this would be expected in a non-steady CFD analysis.

2.5 Evaluation

In the case of Kim et al. (2019) a claim of 700x speed up was made between the ground truth and the surrogate model, however the ground truth was solved on a mid-level CPU and the surrogate model used a high end GPU. This is acknowledged within the paper and the estimated speed up is reduced to a single order of magnitude based of a ratio of memory bandwidth between the CPU and GPU. The paper was distinctly lacking on hard data for MAE between training and test sets. Visually the differences between GT and the surrogate model would be more suitable for, as the paper suggests, quick simulations in games and areas where visual accuracy was of higher importance than absolute accuracy. Where the Kim et al. (2019) paper stood out from the other papers was in the broadness of simulations. Cases involving compressible, incompressible, turbulent, steady-state and non-steady flows could be examined with visually acceptable results.

One notable observation was the similarity between the work of Guo et al. (2016) and Bhatnager et al. (2019) with the later paper obviously following on from the initial work by applying a very similar network design to a more complex problem of aerofoils with more quantification of the aerodynamic behaviour. While the 2D flow field velocity vectors were shown to be accurate within the Bhatnager et al. (2019) paper and a claimed speed up of four orders of magnitude was made on a single GPU in comparison to a RANS solver; there was no solid data provided to back this claim up. It should also be noted that the estimated accuracy was across the entire flow field, of which a significant portion would not be expected to be impacted by minor geometric adjustments. No training or test scores were available or evidence of cross-validation to understand how well the model generalised during training. There was little discussion of hyper-parameter optimisation either.

The results of Wu et al. (2020) showed that low MAE numbers could be achieved for both the test and training set with the training set having a tight distribution for MAE's. Visually it would be difficult for an observer to identify the difference between the surrogate and GT results. However, the test geometries had greater variation that the training set despite being smaller in population [16 samples vs. 500]. Statistically speaking the two groups are different. This backs-up the argument that the network is overfitting to the training data.

The LSTM network trained by Zhang & Zhao (2020) was tested with single and multiple turbines in series. Visually the flow field information appears reasonably accurate, however there is a lack of MAE data to quantify the error. It is claimed there is only a 4.8% error to freestream windspeed.

3. PROFESSIONAL, LEGAL, ETHICAL AND SOCIAL CONSIDERATIONS

In all areas of research special consideration should be given relating to Professional, Legal, Ethical and Social [PLES] issues.

3.1 Professional

The British Computer Society [BCS] sets out a code of conduct for IT professionals, which applies to all members and is split out into four distinct sections:

1. Public Interest
2. Professional Competence and Integrity
3. Duty to the Relevant Authority
4. Duty to the Profession

As researchers in the computer science community it is crucial to be honest about research results, regardless of the actual outcome against the expected. While it might be human nature to want to present a positive outcome, this could bring the research and the profession into dispute (BCS, 2015).

3.2 Legal

One legal consideration in general use of Data Analytics and AI is the misuse of data. The General Data Protection Regulation [GDPR] is not applicable in this instance as the data being processed is neither “Personal” or “Sensitive Personal” (ICO, 2018). However, data used to train and test the models could be confidential and have been shared by industry partners, in confidence, for limited research purposes. This will typically be covered by a Non-Disclosure Agreement [NDA] which should always be consulted before considering publication of said research. NDAs provide protection for two parties interested in sharing information, with the NDA specifically protecting said data (Bott, 2014). It should also be ensured that data provided by an industry partner did not originate from a third-party, such as a customer or client – this is also part of the BCS code of conduct under *Duty to Relevant Authority*.

Another key legal consideration is intellectual property [IP]. This applies not only to the development of models and software, which might infringe copyright law but also how that software might be used to develop new geometries that might themselves be covered under patent law (Bott, 2014). The CFD software market is competitive and it his highly likely that existing CFD software companies are researching in this area in order to maintain a competitive edge. The BCS code of conduct also covers this under *Public Interest* (ICO, 2018).

3.3 Ethical

The research considered in this paper is essentially a study of methods that can take one input of fixed dimensions [such as a 2D image of 250 x 180 pixels] and provides a new output of the same dimensions. While the specific cases here are focussed on emulating CFD, there is no reason the same architecture and methodology could not be applied to photographs or images. Generative ML approaches and specifically VAEs and GANs are powerful tools for image processing (Chollet, 2017) and could be used for nefarious applications. One application would be faking or forging images by merging latent representations from two different sources to generate a new, false, image. A snapshot paper by the Centre for Data Ethics and Innovation [CDEI], published by the UK government advised the research community to support research in detection methods for “DeepFakes” (CDEI, 2019).

4. CONCLUSION, LIMITATIONS & FURTHER WORK

4.1 Conclusion

In steady-state cases it has been found that CNN's can be used as surrogate models in varying architectures to make flow field approximations in place of CFD with reasonable accuracy on simple 2D shapes and aerofoils. CNN's were explored in both autoencoders and using a GAN. In each instance it was found that computation time could be reduced by at least one order of magnitude in comparison to CFD, with claims of near real time flow field predictions once the network was trained. Non-steady cases made use of a LSTM network and provided good approximations, however lacked the same level of fidelity as the steady-state generative models with regards to flow field MAE accuracy.

4.2 Limitations

A key limitation observed is the design space explored by the reviewed models. For instance, the 500 variations of aerofoil on Wu et al. (2020) study were all very similar, with only minor modifications to the profile. If a designer wanted to understand a new concept aerofoil, these models would likely struggle as they would be asked to extrapolate with no data available to direct the models. This would also apply to the output space, whereby the model might quickly need to move to extrapolate an output condition for a given geometry within the original design space.

These papers were in most instances limited to laminar, steady-state flow where CFD would typically be used for more complex turbulent flow conditions. The exceptions were Kim et al. (2019) where a broader range of conditions was examined, with reduced accuracy. Zhang & Zhao (2020) specifically studies a non-steady state problem with a reasonable level of visual accuracy, but limited data to verify with.

4.3 Further Work

This paper has examined how well DNN's can fit flow field data to a specified geometry when compared to CFD. CFD is a physics-based model and is known to have good approximations to hardware experimentation test data. However, there are fields where physics-based CFD models struggle. For instance, radial compressor surge and blade bass noise predictions. Numerical solutions, such as those discussed in this paper, could be advantageous in these fields.

CNN's can be pre-trained with sample latent space representations (Chollet, 2017). Many latent image space datasets exist for image recognition. Developing a similar library to pre-train ROM surrogates for CFD would make sense in order to allow CNNs to generalise better and be more applicable for general use. This could be explored along with understanding which specific features the CNN found to have the highest weighting.

5. REFERENCES

BCS. (2015). *Code of Conduct for BCS Members*. Retrieved 22/10/2020 from https://www.bcs.org/media/2211/bcs-code-of-conduct.pdf

Bender, N. C., Andersen, T. O., & Pedersen, H. C. (2019). Feasibility of Deep Neural Network Surrogate Models in Fluid Dynamics. *Modeling, identification and control, 40*(2), 71-87. https://doi.org/10.4173/mic.2019.2.1

Bhatnagar, S., Afshar, Y., Pan, S., Duraisamy, K., & Kaushik, S. (2019). Prediction of aerodynamic flow fields using convolutional neural networks. *Computational mechanics, 64*(2), 525-545. https://doi.org/10.1007/s00466-019-01740-0

Bott, F. (2014). *Professional Issues in Information Technology*. BCS Learning & Development Limited. http://ebookcentral.proquest.com/lib/hud/detail.action?docID=1511183

CDEI. (2019). *Snapshot Paper - Deepfakes and Audiovisual Disinformation*. gov.uk. Retrieved 21/10/2020 from https://www.gov.uk/government/publications/cdei-publishes-its-first-series-of-three-snapshot-papers-ethical-issues-in-ai/snapshot-paper-deepfakes-and-audiovisual-disinformation#are-deepfakes-a-threat

Chollet, F. (2017). *Deep Learning with Python*. Manning Publications Co.

Gandhi, R. (2018). *Build Your Own Convolution Neural Network in 5 mins.* Retrieved 15/10/2020 from https://towardsdatascience.com/build-your-own-convolution-neural-network-in-5-mins-4217c2cf964f

Guo, X., Li, W., & Iorio, F. (2016). *Convolutional Neural Networks for Steady Flow Approximation* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA. https://doi.org/10.1145/2939672.2939738

ICO. (2018). *Guide to the General Data Protection Regulation*. Retrieved 21/10/2020 from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/711097/guide-to-the-general-data-protection-regulation-gdpr-1-0.pdf

Kim, B., Azevedo, V. C., Thuerey, N., Kim, T., Gross, M., & Solenthaler, B. (2019). Deep Fluids: A Generative Network for Parameterized Fluid Simulations. *Computer graphics forum, 38*(2), 59-70. https://doi.org/10.1111/cgf.13619

Li, K., Kou, J., & Zhang, W. (2019). Deep neural network for unsteady aerodynamic and aeroelastic modeling across multiple Mach numbers. *Nonlinear dynamics, 96*(3), 2157-2177. https://doi.org/10.1007/s11071-019-04915-9

Mukesh, R., Mukesh, R., Lingadurai, K., Lingadurai, K., Selvakumar, U., & Selvakumar, U. (2018). Airfoil Shape Optimization based on Surrogate Model. *Journal of The Institution of Engineers (India): Series C, 99*(1), 1-8. https://doi.org/10.1007/s40032-017-0382-x

Tao, J., & Sun, G. (2019). Application of deep learning based multi-fidelity surrogate model to robust aerodynamic design optimization. *Aerospace science and technology, 92*, 722-737. https://doi.org/10.1016/j.ast.2019.07.002

Wu, H., Liu, X., An, W., Chen, S., & Lyu, H. (2020). A deep learning approach for efficiently and accurately evaluating the flow field of supercritical airfoils. *Computers & fluids, 198*, 104393. https://doi.org/10.1016/j.compfluid.2019.104393

Zhang, J., & Zhao, X. (2020). A novel dynamic wind farm wake model based on deep learning. *Applied energy, 277*, 115552. https://doi.org/10.1016/j.apenergy.2020.115552