A surrogate CFD model using Machine Learning for fast design explorations of the indoor environment

Antonio D’Aquilio1, Hesham Ebrahim1, Enrico Crobu1
1Hoare Lea Llp, London, United Kingdom

Abstract
The use of Computational Fluid Dynamics (CFD) to evaluate indoor air quality and thermal comfort performance has been widely adopted. However, it remains a computationally demanding process that requires prior experience in commercially available codes. Thus, it cannot be used adequately in a fast-paced early HVAC design stage. This paper explores the efficacy of a surrogate model for real-time prediction of steady turbulent flow within the context of the indoor environment. The proposed approach employs Convolutional Neural Networks (CNNs) to predict physical properties of the environment such as air velocity, dry-bulb temperature and age of air on a plane positioned at 1.2m above ground. The proposed surrogate model is capable of predicting results with low mean average errors (0.315, 0.931, 0.017, 0.016 for temperature, age of air, Ux and Uy respectively) and four orders of magnitude faster than a standard CFD simulation for a similar scenario, thereby enabling fast explorations of design options, especially useful in the early stage of the design.

Highlights
- Convolutional Neural Networks are able to learn spatial relationships between geometric inputs and physical variables.
- Heating, Ventilation and Air Conditioning design optimisation can benefit from Machine Learning techniques.
- Low-error surrogate models can enable fast exploration of design options at the early stages.

Introduction
The evaluation of the indoor environment performance has gained increasing relevance in recent years, primarily due to the need to mitigate the effects of climate change (Coley et al., 2020). The building industry has expanded the scope of performance by recognising the impact of the indoor environments on occupants’ productivity over the past decade (Kawakubo et al., 2023). This led to the development of new indoor comfort standards that consider both comfort and productivity (Ildiri et al., 2022).

Computational Fluid Dynamics (CFD) is a numerical method used to solve the Navier-Stokes equations that provides information about the airflow distribution pattern and physics based on a given set of boundary conditions. Environmental consultants and researchers have widely adopted the use of CFD to investigate the indoor environment and HVAC design over the last two decades (Morozova et al., 2020; Chen et al., 2004). The detailed spatial indoor analysis and its impact on the energy performance of HVAC design has also been investigated using CFD techniques (Dai et al., 2022). Despite its effectiveness in yielding promising predictions, even for steady-state simulations, CFD is regarded as computationally expensive, making it unsuitable for early-stage HVAC and architectural design that rapidly evolve. Therefore, to achieve optimal design outcomes, it is vital to use quick performance evaluation tools that can be used to optimise the indoor environment with regards to indoor comfort and air quality.

Machine Learning (ML) techniques, such as Deep Learning (DL) and Convolutional Neural Network (CNN), have been proven to be successful at learning complex, high dimensional and non-linear relationships from large datasets of examples (LeCun et al., 1995). The use of ML techniques for CFD analysis has been extensively studied over the last decade (Calzolari et al., 2021). Previous studies have explored several ML approaches for predicting the aerodynamics of solid objects (Usman et al., 2021) and buildings (Musil et al., 2019; He et al., 2021), as well as their use in predicting indoor air quality (Buratti et al., 2020) and optimising HVAC system controls (Tien et al., 2022; Shin et al., 2023). In recent years, research has investigated the application of DL techniques for the indoor airflow prediction (Zhou et al., 2021). A notable advancement in CNN architecture is U-net, introduced by Ronneberger et al. (2015) for effective image-to-image mapping in biomedical image segmentation. Ribeiro et al. (2020) proposed a variant of the U-net architecture, which the authors used to predict laminar air flow properties (such as pressure and air velocity), around arbitrary object shapes in a wind tunnel setup. This research builds on Ribeiro et al.’s application and adapts their methodology to examine the feasibility and limitations of using ML for real-time performance evaluation of the indoor HVAC design. The goal is to facilitate discussions regarding optimisation of the airflow distribution during early stages of the design process. To achieve this, the paper focuses...
on the physics affecting two key aspects of the indoor environment performance, namely thermal comfort and indoor air quality. The proposed model utilises a CNN approach to learn from validated indoor CFD analysis data and predict air properties, including temperature, air velocity and age of air, based on a set of geometrical boundary conditions. The model can discern patterns and non-linear relationships between geometric inputs and airflow characteristics. With relatively low error margins, this application can be valuable for optimising a HVAC design layout with regards to indoor comfort and air quality.

The proposed methodology is tailored to indoor spaces that are fully controlled via a HVAC system for a well-mixed indoor temperature environment. It is particularly applicable to commercial spaces such as office, retail and laboratories etc. The surrogate model is trained on a 2D surface located at 1.20m, hence HVAC systems that rely on air temperature stratification, such as Underfloor Air Distribution, may require further feature mappings. For instance, incorporating variables like floor-to-ceiling height would fall beyond the scope of the present study.

Methodology

Geometry

To generate the training data, a parametric room was setup to replicate an open plan office space. The office space is of fixed dimensions of 20 m wide, 12 m deep and 4 m high. The room has been setup with supply and extract grilles for ventilation and conditioning. Internal gains (lighting, people, equipment) have been distributed homogenously within the room, as illustrated in Figure 1.

For simplicity, the internal gains were modelled as box-like volumes. The lighting gains were located at 0.2 meters from the ceiling surface. The people gains (representing people sitting at a desk) were located at 0.50 m from the floor, while the equipment gains (representing computer screens) were located at 1.50 m from the floor. Further, the supply and extract grilles were located on the ceiling of the room. The presence of an external façade was also included to account for localised convective gains emitted by the surface. However, the location and size of the façade were not varied for this experiment. Finally, the internal room layout does not take into account presence of other internal obstructions typical of an office (e.g. desks, chairs, cabinets etc.). As shown in Figure 2, the room was parametrically setup to generate variations of the internal layout with regards to:

- Supply grilles number and disposition
- Extract grilles number and disposition
- Internal walls location and size

By varying the above parameters, the room’s internal layout results in different distributions of internal gains.

Figure 2: Plan view of the parametric room setup. a) base-setup, b) change in extract grilles number and disposition, c) change in supply grilles number and disposition, d) change in location and size of internal walls.

The room setup was used to generate OpenFOAM case folders and parametrically run different internal layout configurations.

CFD Setup

Computational Fluid Dynamics (CFD) simulations were conducted to investigate heat transfer phenomena in a room using OpenFOAM (v9). The buoyantSimpleFoam solver was utilised to model the fluid dynamics and heat transfer within the room. The simulation domain was discretised using a hexahedral grid topology generated with snappyHexMesh, resulting in a mesh with approximately 2 million cells (see Figure 3). To smooth the buoyancy-driven spikes of the CFD results spatially, the solution was averaged over the last 500 iterations.

Figure 3: Hexahedral grid topology of the room.

The setup of the boundary conditions for the simulations follows the configuration illustrated in Figure 1. The heat gains from the equipment, people, lighting, and façade are held constant across all simulations, as are the supply temperatures and velocity. Further information regarding these boundary conditions is provided in Table 1.
The adopted settings of the simulations are summarised in Table 2.

Table 2: Adopted simulation settings.

<table>
<thead>
<tr>
<th>Modelling Parameters</th>
<th>Adopted Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grid Topology</td>
<td>Hexahedral mesh</td>
</tr>
<tr>
<td>Number of Cells</td>
<td>~11 million</td>
</tr>
<tr>
<td>Prism Layer Count</td>
<td>3 cells</td>
</tr>
<tr>
<td>Time</td>
<td>Steady State</td>
</tr>
<tr>
<td>Equation of State</td>
<td>Boussinesq - $\rho = 1.225$</td>
</tr>
<tr>
<td>RANS Model</td>
<td>$k-\varepsilon$ realizable</td>
</tr>
<tr>
<td>Transport Model</td>
<td>Newtonian (Air)</td>
</tr>
<tr>
<td></td>
<td>$\nu = 1.56e-5 , m^2/s$</td>
</tr>
<tr>
<td></td>
<td>$\beta = 3.35e-3$</td>
</tr>
<tr>
<td></td>
<td>$Pr = 0.7$</td>
</tr>
<tr>
<td></td>
<td>$Pr_t = 0.7$</td>
</tr>
<tr>
<td></td>
<td>$C_{\rho 0} = 1006$</td>
</tr>
<tr>
<td>Numerical Schemes</td>
<td>Second</td>
</tr>
</tbody>
</table>

Where $\rho$ is the density of air $\nu$ is the laminar kinematic viscosity of air, $\beta$ is the thermal expansion coefficient, $Pr$ is the Laminar Prandtl number and $Pr_t$ is the turbulent Prandtl number and $C_{\rho 0}$ is the specific heat of air.

CFD Validation

To verify the simulation approach and ensure sufficiently accurate training data for the surrogate model, the results were compared to ANSYS Fluent, a well-established commercial solver, using the same mesh and physics due to the lack of experimental investigation of the proposed geometrical setup.

The comparison of results was primarily based on the mean temperature and velocity across a horizontal plane (in green) and a vertical plane (in orange) positioned 1.2 m above ground and at the centreline, respectively, as shown in Figure 4.

The dry-bulb temperature results of the horizontal plane presented in Figure 5 indicate a generally uniform temperature range of 24-26°C, with localised higher temperatures surrounding the people gains, as expected. Interestingly, the temperature distribution differs by 1°C between the two CFD codes, as Fluent shows a lower temperature compared to OpenFOAM. Additionally, minor differences are observed in the interaction of hot spots between the people gains.

Analysis of the dry-bulb temperature from the vertical plane (Figure 6) reveals that buoyancy drives the temperature variation in the room, with hotter air being saturated at the top and cooler air at the bottom. Both CFD codes appear to provide similar qualitative temperature distributions, although Fluent estimates the temperature to be slightly cooler than OpenFOAM up to the mid-height of the room. Minor differences are also observed in the jetted flow and its interaction with the people and equipment on the furthest supply to the eastern side of the room.

![Figure 4: The planes used to sample the results of the room for qualitative analyses.](image)

![Figure 5: Dry-bulb temperature results of the horizontal plane 1.2 m above ground. OpenFOAM results at the top and Fluent at the bottom.](image)

Proceedings of the 18th IBPSA Conference
Shanghai, China, Sept. 4-6, 2023

https://doi.org/10.26868/25222708.2023.1590
When comparing the velocity field across the horizontal plane, eight distinctively circular higher velocity zones are observed, which represent the jetted flow from the supplies at the roof, as presented in Figure 7. Notably, the location and magnitude of these velocities correlate reasonably well between the two CFD codes. However, differences are observed in the vortex structures and distribution of slower velocities at the perimeter of the room. This behaviour is also evident in the vertical plane shown in Figure 8, where the jetted flow forms rotating eddies similar to those seen in vertical take-off and landing aircraft (Page et al., 2005).

The histogram in Figure 9 was plotted to better understand the absolute difference in dry-bulb temperature between the two CFD codes. It reveals that there is a difference of 1°C between the CFD codes over the highest number of cells.

The variations observed in the results could be ascribed to differences in the numerical schemes used for discretisation, the equations coupling pressure and velocity, the treatment of near-wall regions and the specific implementation of boundary conditions, all of which differ between the two CFD codes. Nonetheless, the results obtained are consistent with anticipated physical phenomena and heat balance equations, and thus provide a reasonable foundation for training the surrogate model, which is the main objective of this investigation.

**CFD Analysis Post-process**

The CFD analysis results were post-processed to extract data based on a uniform 2D grid with a 0.20 m cell size located at 1.20 m from the floor. Forty simulations were carried out, and a simple data augmentation technique was used to increase the number of training samples. The CFD analysis results were mirrored vertically and horizontally, and in both directions to increase the training samples by a factor of four. Thus, the total dataset consisted of 160 samples. The dataset was split with a 70-30% ratio between training and test datasets. The training set was used to train the model, while the test dataset was used to evaluate its performance.
Proposed CNN Architecture
The adopted CNN architecture is based on the U-net variant proposed by Ribeiro et al. (2020). However, the model has been adapted to enable the mapping of the relationships between geometric information and different spatial physics outputs associated with the indoor environment and HVAC design.

The proposed architecture, as illustrated in Figure 10, consists of one encoder and four decoders. The model makes use of 2D input data and results in 2D outputs. Five input channels were set to encode spatial information with regards to:

- External walls
- Internal walls
- Supply grilles
- Return grilles
- Multi-class channel

The encoder specialises in extracting 2D input information and encodes it into a Latent Geometry Representation (LGR). The LGR is then decoded and upsampled back to match the input image size via four decoders. Each decoder specialises in the output of different physical properties, namely:

- Dry-bulb air temperature
- Age of air
- Ux (x component of the velocity vector)
- Uy (y component of the velocity vector)

The distance data was based on the Signed Distance Function (SDF) proposed by Guo et al. (2016). However, it was noted that using negative values for the distance function reduced the model’s performance. Thus, the negative component of the signed distance function was ignored, keeping 0 values for the grid points located within an object (internal walls, supply/extract grilles). The multi-class channel contains regional information of the sampled grid, with 0 used to denote air, 1 for external wall boundary and 2 for internal walls.

The loss function used for the training of the proposed model is defined as follows.

\[
L = \frac{1}{n} \sum_{i=1}^{n} |y_i^T - \hat{y}_i^T| + \frac{1}{n} \sum_{i=1}^{n} |y_i^{AoA} - \hat{y}_i^{AoA}| + \\
\frac{1}{n} \sum_{i=1}^{n} (y_i^{Ux} - \hat{y}_i^{Ux})^2 + \frac{1}{n} \sum_{i=1}^{n} (y_i^{Uy} - \hat{y}_i^{Uy})^2
\]  

Where:

- n: the number of data samples in the batch
- \( y_i^T \): the ith-element of the dry-bulb temperature ground truth
- \( \hat{y}_i^T \): the ith-element of the dry-bulb temperature prediction of the model
- \( y_i^{AoA} \): the ith-element of the age of air ground truth
- \( \hat{y}_i^{AoA} \): the ith-element of the age of air prediction of the model
- \( y_i^{Ux} \): the ith-element of the x component of the velocity vector ground truth
- \( \hat{y}_i^{Ux} \): the ith-element of the x component of the velocity vector prediction
- \( y_i^{Uy} \): the ith-element of the y component of the velocity vector ground truth
- \( \hat{y}_i^{Uy} \): the ith-element of the y component of the velocity vector prediction

The loss function is responsible for aggregating the losses for all field outputs. For the two velocity components, the Mean Squared Error (MSE) was used as the loss function, while the Mean Absolute Error (MAE) was used for dry-bulb temperature and age of air. The decision to use distinct loss components for the model outputs was based on preliminary evaluations, which demonstrated superior performance gains compared to applying a uniform MSE to all loss components.

The loss is computed for each data batch, and the gradient of the loss (with respect to the network’s weights) is computed and back-propagated through the network for each batch, to optimise the model and reduce the loss. The loss is then accumulated for all batches and averaged on each epoch for visualisation purposes.

Model Training
Initial model testing has been carried out to define baseline performance and to allow for further optimisation. The model was found to effectively learn by allowing for 500 epochs, with a learning rate set to 0.001. The AdamW optimizer was utilised to minimize the network’s loss. For consistency, the learning rate and number of epochs were fixed during the hyperparameter
tuning. There is scope for further investigation over the impact of these parameters on overall performance.

Performance metrics were defined to evaluate the model effectiveness at predicting the output physics fields. Specifically, the MAE was adopted for each of the four output variables (dry-bulb temperature, age of air, Ux and Uy). The MAE metric allows for a more intuitive comparison of the results with units aligned to the outputs variable’s units. This metric is computed and reported at each epoch. During the 500 epochs of training, the model with the lowest error loss is saved and its metrics are reported.

**Hyperparameter Tuning**

A grid-search approach has been adopted to explore the model performance based on its hyperparameters configuration. The tested hyperparameters are summarised in Tables 3 and 4.

**Table 3: Hyperparameters tested.**

<table>
<thead>
<tr>
<th>Batch-weight norm</th>
<th>Kernel size</th>
<th>Weight decay</th>
<th>Filters</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-On</td>
<td>3</td>
<td>0.005</td>
<td>8-16-32-32</td>
</tr>
<tr>
<td>Off-Off</td>
<td>5</td>
<td>0.1</td>
<td>8-16-32-64</td>
</tr>
</tbody>
</table>

**Table 4: Hyperparameters tested (continued).**

<table>
<thead>
<tr>
<th>Batch-weight norm</th>
<th>Kernel size</th>
<th>Weight decay</th>
<th>Filters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off-Off</td>
<td>3</td>
<td>0.005</td>
<td>8-16-32</td>
</tr>
<tr>
<td>Off-Off</td>
<td>5</td>
<td>0.1</td>
<td>4-8-16-32</td>
</tr>
</tbody>
</table>

The selected hyperparameter options resulted in 32 combination scenarios. Important hyperparameters of a CNN model are its width and depth. The width refers to the number of filters that the model user for feature mapping, which allows it to extract a larger number of features from the input data. The depth of the model refers to the number of convolutional downsampling, which allows for higher level of non-linear abstraction and can increase the model’s complexity. These two aspects have been captured during the hyperparameter tuning by varying the number of filters and the number of convolutional downsampling steps. The kernel size impacts the feature map size, which is an important aspect with regards to the encoded geometry information extracted from the input data. With smaller feature maps, the model is able to capture fine-grained geometry relationships, while potentially suffering from the loss of sight for larger geometry relationships. An effective technique to reduce overfitting within DL architectures is the use of regularization terms additional to the loss function. For this purpose, different weight decays have been tested to understand the impact of the regularization term implemented by the optimizer.

**Results**

This section of the paper analyses the outcomes of model training and hyperparameter tuning. A performance review has been carried out focusing on specific test metrics. A qualitative review of spatial outputs was also conducted to assess the level of spatial accuracy.

**Performance Review**

Table 5 summarises the top five best performing options of hyperparameters (and resulting model) based on the test dataset. The MAE results of the best performing model for dry-bulb temperature, age of air, Ux and Uy are 0.315, 0.931, 0.017 and 0.016, respectively.

**Table 5: Test metrics of best performing models.**

<table>
<thead>
<tr>
<th>#</th>
<th>MAE Temperature</th>
<th>MAE Age of air</th>
<th>MAE Ux</th>
<th>MAE Uy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.315</td>
<td>0.931</td>
<td>0.017</td>
<td>0.016</td>
</tr>
<tr>
<td>2</td>
<td>0.318</td>
<td>0.927</td>
<td>0.018</td>
<td>0.018</td>
</tr>
<tr>
<td>3</td>
<td>0.324</td>
<td>0.938</td>
<td>0.018</td>
<td>0.017</td>
</tr>
<tr>
<td>4</td>
<td>0.325</td>
<td>0.965</td>
<td>0.018</td>
<td>0.019</td>
</tr>
<tr>
<td>5</td>
<td>0.331</td>
<td>0.967</td>
<td>0.018</td>
<td>0.017</td>
</tr>
</tbody>
</table>

**Table 6: Best performing models hyperparameters.**

<table>
<thead>
<tr>
<th>#</th>
<th>Batch-weight norm</th>
<th>Kernel size</th>
<th>Weight decay</th>
<th>Filters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>On-On</td>
<td>5</td>
<td>0.010</td>
<td>8-16-32-64</td>
</tr>
<tr>
<td>2</td>
<td>On-On</td>
<td>5</td>
<td>0.005</td>
<td>8-16-32-32</td>
</tr>
<tr>
<td>3</td>
<td>On-On</td>
<td>5</td>
<td>0.010</td>
<td>8-16-32-32</td>
</tr>
<tr>
<td>4</td>
<td>On-On</td>
<td>3</td>
<td>0.005</td>
<td>8-16-32-64</td>
</tr>
<tr>
<td>5</td>
<td>On-On</td>
<td>5</td>
<td>0.005</td>
<td>8-16-32-32</td>
</tr>
</tbody>
</table>

As shown in Table 6, the best performing model uses a deep network with four downsampling steps with increasing number of feature maps, a kernel size of 5 and a weight decay set at 0.01. The option with an even deeper architecture (five downsampling steps) did not make the top result, suggesting that there may be a loss of spatial information when downsampling occurs five times when compared to four. Moreover, the best performing model uses a higher weight decay parameter, which suggests that weight decay can help the model generalise better and reduce overfitting.

The worst performing model was also illustrated in Tables 5 and 6. It is notable that the five best performing options had both batch and weight normalisation switched on, while the lowest performing option had neither. Figures 12 and 13 depict the training and test loss curves for the models, demonstrating that batch and weight normalisation leads to faster model convergence.
This confirms that the batch and weight normalisation can accelerate model learning for this dataset. Further investigations are required to identify whether the models without normalization would require further training epochs to ensure a more equitable comparison.

**Qualitative Review**

In this section, the best performing model is used to run predictions on a set of geometric inputs belonging to the test dataset (i.e. on ‘unseen’ data). Figures 14 and 15 illustrate the inference results for two randomly selected cases from the test dataset.

The performance evaluation of the model on unseen data reveals its capability to generalise and predict the spatial distribution of temperature, age of air and velocity components. It successfully predicts the changes in overall temperature and its distribution due to the variations in boundary conditions, allowing the identification of potential hotspots. The proposed model is able to also predict the age of air pattern relatively well, which can aid in identifying areas of air stagnation and pollutant accumulation during the early stages of HVAC design. The velocity components fields yield relatively close results, with areas of acceleration generally matching the CFD results.

When comparing the results for the same test cases for the worst-performing model, as shown in Figure 15, it can be observed that the model is still able to predict the age of air and velocity components relatively well. However, the temperature field shows temperature variations that deviate greatly from the CFD ground truth. Although, as previously noted, this model may require further training to enable a fairer comparison. The difference between the two models may indicate that the deeper architecture with a larger number of filters can help the model abstract non-linear relationships between geometric inputs and resulting physics variables.

**Conclusions and Future Work**

This paper illustrated a CNN model capable of predicting the spatial distribution of physics variables based on a set of geometric boundary conditions. The MAE results of the proposed model for dry-bulb temperature, age of air, Ux and Uy are 0.315, 0.931, 0.017 and 0.016, respectively. Moreover, the proposed CNN model can produce insightful outputs four orders of magnitude faster than a standard CFD approach. For the testing done in this...
research, the prediction time of the CNN model was <1s, compared to 75 minutes for the CFD simulation. The surrogate model’s high speed and low error margin make it a promising tool for enhancing the HVAC layout and configuration during early design phases, particularly with regards to thermal comfort and air quality.

The use of open-source CFD code to generate extensive training data can be a valuable approach when compared to expensive commercial software alternatives. Further research is needed to expand the model’s scope to encompass a broader range of variables, including different room boundaries and internal configurations, a wide range of supply velocities and temperatures, and a greater variety of architectural features (such as taller volumes and different envelope conditions). Finally, the applicability of the proposed architecture for 3D predictions will also be investigated.

References


