# Super-Resolution of Three-Dimensional Temperature and Velocity for Building-Resolving Urban Micrometeorology Using Physics-Guided Convolutional Neural Networks with Image Inpainting Techniques

Yuki Yasuda<sup>a,1</sup>, Ryo Onishi<sup>a</sup>, Keigo Matsuda<sup>b</sup>

 <sup>a</sup>Global Scientific Information and Computing Center, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, 1528550, Japan
 <sup>b</sup>Research Institute for Value-Added-Information Generation, Japan Agency for Marine-Earth Science and Technology, 3173-25 Showa-machi, Kanazawa-ku, Yokohama, Kanagawa, 2360001, Japan

# Abstract

Atmospheric simulations for urban cities can be computationally intensive because of the need for high spatial resolution, such as a few meters, to accurately represent buildings and streets. Deep learning has recently gained attention across various physical sciences for its potential to reduce computational cost. Super-resolution is one such technique that enhances the resolution of data. This paper proposes a convolutional neural network (CNN) that super-resolves instantaneous snapshots of three-dimensional air temperature and wind velocity fields for urban micrometeorology. This superresolution process requires not only an increase in spatial resolution but also the restoration of missing data caused by the difference in the building shapes that depend on the resolution. The proposed CNN incorporates gated convolution, which is an image inpainting technique that infers missing pixels.

*Email addresses:* yasuda.y.aa@m.titech.ac.jp (Yuki Yasuda),

onishi.ryo@gsic.titech.ac.jp (Ryo Onishi), k.matsuda@jamstec.go.jp (Keigo Matsuda)

<sup>&</sup>lt;sup>1</sup>Corresponding author. 2-12-1-i9-111 Ookayama, Meguro-ku, Tokyo, 1528550, Japan. Tel.: +81 3 57343173 (office). Email address: yasuda.y.aa@m.titech.ac.jp

The CNN performance has been verified via supervised learning utilizing building-resolving micrometeorological simulations around Tokyo Station in Japan. The CNN successfully reconstructed the temperature and velocity fields around the high-resolution buildings, despite the missing data at lower altitudes due to the coarseness of the low-resolution buildings. This result implies that near-surface flows can be inferred from flows above buildings. This hypothesis was assessed via numerical experiments where all input values below a certain height were made missing. This research suggests the possibility that building-resolving micrometeorological simulations become more practical for urban cities with the aid of neural networks that enhance computational efficiency.

*Keywords:* super-resolution, image inpainting, convolutional neural network, building-resolving micrometeorological model, large eddy simulation

# 1. Introduction

Global urbanization indicates that various challenges in cities, such as reducing heat stress and energy consumption, will become more important in the near future [1]. The estimation of environmental conditions such as wind and temperature may be useful for addressing these tasks [2, 3]. Indeed, architects and urban planners are now incorporating the knowledge of urban micrometeorology, e.g., thermal and dynamical airflow responses, into building designs [4–6]. Numerical simulation based on the laws of physics is an effective tool for estimating atmospheric states on micrometeorological scales [5, 6]. However, these simulations can be computationally intensive due to the need for high spatial resolution such as a few meters to represent buildings and streets. Therefore, the reduction in computational cost is necessary to make micrometeorological simulations more practical for various applications in urban areas.

Recently, super-resolution (SR) has been proposed as a promising ap-

proach for reducing computational cost in numerical simulations across various physical sciences [e.g., 7–9]. Super-resolution is a technique that enhances the resolution of images and has been focused on in computer vision as an application of neural networks (NNs) [10–12]. The success of such NNs has led to numerous applications of SR for various fluid systems [13–24], including atmospheric flows [7, 25–30]. The recent progress in the fluid SR is summarized in a review paper [31]. One application in meteorology is SR simulation [7, 25, 26], where the time evolution is calculated from low-resolution (LR) atmospheric simulations, and high-resolution (HR) inferences are obtained by inputting the LR results into a trained NN. SR simulation is computationally efficient because it only requires LR atmospheric simulations and the elapsed time of NN inference is usually negligible.

The potential of SR simulation has been investigated for urban cities [7, 25, 30]. These studies focused on the SR of two-dimensional fields such as temperature. They reported that the multiple input improves the validity of inference because NNs can learn the physical relationships between these variables [25, 30]. For more practical applications, however, it is necessary to super-resolve three-dimensional fields such as wind velocity.

It may be challenging to super-resolve micrometeorological simulation results in cities because the streets between buildings may not be resolved at LR. In such simulations, the atmospheric flows outside buildings are represented, while the grid points inside buildings are missing. Thus, the SR process requires not only an increase in spatial resolution but also the reconstruction of flows around buildings to account for changes in building shape from LR to HR. Our previous studies on the two-dimensional SR [7, 30] avoided this issue by sampling temperature at a certain height above the bottom surface such as the ground or roofs of buildings. As a result of this sampling, the two-dimensional data contained no missing value. Furthermore, in other fluid systems, such as flows around airfoils [18–20] and cylinders [15–18], the changes in obstacle shape during the SR process have not been considered. Therefore, it is currently unclear how to super-resolve flows around buildings whose shape depends on the resolution.

In computer vision, the reconstruction of missing pixels is referred to as image inpainting, and deep-learning based methods have been studied [32, 33]. For instance, missing pixels are considered masked and are recovered based on the surrounding pixels or statistical information learned from data. These techniques have been applied to various observation data, including satellite imagery, to fill in missing values caused by various factors such as cloud cover, measurement error, or data sparsity [34–40]. However, these methods cannot be directly applied to urban micrometeorological data because missing regions remain after the conversion from LR to HR. In other words, HR atmospheric flows are still considered masked by HR buildings.

The changes in the shape of missing regions may be addressed by introducing two masks at LR and HR. However, there are few studies utilizing multiple masks. Schweri et al. [39] developed a U-Net [41] by using two types of masks to reconstruct the atmospheric vortex street, which is missed due to cloud cover, behind Guadalupe Island. The first mask, called the obstacle mask, represents obstacles (i.e., Guadalupe Island) that remain in the output. The other mask, called the inpainting mask, corresponds to cloud cover and is varied in partial convolution [42]. The inpainting mask takes 0 at missing pixels and 1 at non-missing pixels. In the partial convolution [42], missing pixels are restored by convolution based only on the surrounding non-missing pixels, while the inpainting mask is shrunk according to this restoration. Although the U-Net developed by Schweri et al. [39] successfully restored missing data of horizontal flows around obstacles, they did not perform SR or consider three-dimensional flows. To the authors' knowledge, the combination of SR and image inpainting has yet to be investigated for turbulent flows in three dimensions.

This paper proposes a convolutional neural network (CNN) that superresolves three-dimensional air temperature and wind velocity fields in urban cities. The SR process in the CNN accounts for the building-shape difference between LR and HR. This CNN employs an image inpainting technique called gated convolution [43], which is an extension of the partial convolution [42]. The CNN performance was evaluated by utilizing building-resolving micrometeorological simulations around Tokyo Station in Japan. Specifically, the LR input data were created from the HR simulation results and used to train the CNN via supervised learning. The training loss function included a physics loss term, namely the mean divergence loss, to improve the validity of inference. This proof of concept demonstrates the feasibility of enhancing spatial resolution while simultaneously reconstructing flows around obstacles. This is an essential step toward the development of SR simulation for urban micrometeorological predictions.

The remaining paper is organized as follows. Section 2 presents the CNN for the SR of atmospheric flows around buildings. Section 3 describes the methods to train and evaluate the CNN. The results are analyzed in Section 4. We discuss the possibility of qualitatively inferring near-surface flows from upper flows. Finally, the conclusions are given in Section 5.

# 2. CNN for super-resolution accounting for building-shape changes

The proposed CNN enhances the spatial resolution of three-dimensional air temperature and wind velocity fields on Cartesian coordinates, while simultaneously modifying the missing regions associated with changes in building shape from LR to HR. The input consists of five quantities: LR temperature  $T^{\text{LR}}$  and velocity  $(u^{\text{LR}}, v^{\text{LR}}, w^{\text{LR}})$ , together with HR building mask  $B^{\text{HR}}$ . The output is composed of four quantities: HR temperature  $T^{\text{HR}}$  and velocity  $(u^{\text{HR}}, v^{\text{HR}}, w^{\text{HR}})$ . Here, the eastward, northward, and upward components of the velocity vector V are represented by u, v, and w, respectively. In the following, the superscripts of LR and HR are omitted if there is no confusion. The temperature and velocity are missing at the grid points inside buildings. All input and output are instantaneous snapshots and are described as three-dimensional numerical arrays that correspond to the geometrical distributions in the three-dimensional space.

The building mask B is a binary field that takes 1 and 0 at the grid points outside and inside buildings, respectively:

$$B = \begin{cases} 1 & \text{(outside buildings)}, \\ 0 & \text{(inside buildings)}. \end{cases}$$
(1)

We assume that buildings are described by voxels and each grid point is located in the center of the voxel, i.e., it is not on building walls or roofs. Further, the building mask B is assumed to be static. By element-wise multiplication, B can extract non-missing values outside buildings from any field quantity. Only the HR building mask is utilized, and the LR mask is not required, as explained below.

Figure 1 shows the network architecture of the proposed CNN, which is based on the U-Net developed by Schweri et al. for image inpainting [39]. Hyperparameters such as kernel size were tuned through a grid search. The detailed implementation, including the hyperparameters, can be found on the Zenodo repository (see Data availability). The calculation process in the CNN consists of four steps. First, the LR input of  $T^{\text{LR}}$ ,  $u^{\text{LR}}$ ,  $v^{\text{LR}}$ , and  $w^{\text{LR}}$  is aligned in size with the HR output by using the nearest neighbor interpolation. Second, the resized input is downsampled with gated convolution [43], which is an image inpainting technique described in the next paragraph. In each downsampling block, the first gated convolution has a stride of  $2 \times 2 \times 2$ , halving the array size in each dimension. Third, the downsampled input is transformed by ordinary convolution and Leaky ReLU [44]. Fourth, the upsampling blocks restore the array size to the HR. In each upsampling block, the output from the corresponding downsampling block is passed through a skip connection [41, 45]. Upsampling is then performed by using the threedimensional version of pixel shuffle [46], where the array size is doubled in each dimension. Importantly, the HR building mask is resized with average

pooling where necessary and fed to each downsampling or upsampling block [39].



Figure 1: Architecture of the convolutional neural network (CNN) proposed in this study. Details of Nonlinear Mapper, Upsampling Block, Downsampling Block, and Gated Conv3D are displayed in the lower panels. The label "AvgPool3D" is three-dimensional average pooling, "Conv3D" is three-dimensional convolution, "LkyReLU" is Leaky ReLU [44], "Shuffle" means the three-dimensional version of pixel shuffle [46], and  $\otimes$  represents element-wise multiplication. If two or three arrows point to a block, all inputs are concatenated and then fed to the first layer in the block. The detailed implementation is available on the Zenodo repository (see Data availability).

In the gated convolution [43], the features are multiplied with the weights calculated by a sigmoid function (Fig. 1). This operation can be viewed as an extension of the partial convolution [42]. The partial convolution is rulebased, where only non-missing values are convolved using a prescribed mask. In contrast, the gated convolution is expected to optimize the mask, namely the weights for non-missing values, depending on the input via the sigmoid function. The following preprocessing is applied to regard zeros as missing values. Each variable X (=  $T^{\text{LR}}$ ,  $u^{\text{LR}}$ ,  $v^{\text{LR}}$ , or  $w^{\text{LR}}$ ) is transformed as

$$\operatorname{clip}_{[0,1]}\left(\frac{X-m_X}{s_X}\right).$$
(2)

The clipping function,  $\operatorname{clip}_{[0,1]}(x) = \min\{1, \max\{0, x\}\}\ (x \in \mathbb{R})$ , limits the value range to [0, 1]. The parameters  $m_X$  and  $s_X$  are determined such that 99.9% of X values are within the range of [0, 1]. After applying (2), NaN (Not a Number) values at the grid points inside buildings are replaced with zeros. Due to this preprocessing, the spatial distribution of zeros in the input represents the LR buildings, and the LR building mask is not required.

## 3. Methods

### 3.1. Building-resolving micrometeorological simulations

Building-resolving micrometeorological simulations were conducted by employing the Multi-Scale Simulator for the Geoenvironment (MSSG) [47– 50]. MSSG is a coupled atmosphere-ocean model that operates on global, meso-, and urban scales. The atmospheric component of MSSG, i.e., MSSG-A, can be utilized as a building-resolving large eddy simulation (LES) model in conjunction with a three-dimensional radiative transfer model [50]. The governing equations of MSSG-A include the conservation equations of mass, momentum, and energy for compressible flows and the transport equations for mixing ratios of water substances including water vapor, liquid, and ice cloud particles. Further information on the configuration can be found in our precursor studies [7, 50].

Dynamical downscaling [51] was performed for an actual urban area around Tokyo Station in Japan (35.6809°N and 139.7670°E) to reduce the spatio-temporal scales of atmospheric flows from meso- to urban scales. The mesoscale simulations used three two-way-coupled nested systems (Fig. 2): Domain 1 through 3. These domains are characterized by horizontal grid points of  $160 \times 160$ , with grid spacings of 1 km for Domain 1, 300 m for Domain 2, and 100 m for Domain 3. The vertical grids for Domains 1 through 3 are identical and consist of 55 points over an altitude range of 40 km. The computation domain of micrometeorology, i.e., Domain Tokyo (Fig. 2), is nested within Domain 3. This smallest domain has horizontal grid points of  $400 \times 400$  with a resolution of 5 m, as well as 151 stretched grid points in the vertical dimension. The vertical grid spacings are 5 m below an altitude of 500 m and varied between 500 and 1500 m. The building height distribution was created from the geographic information system (GIS) data for 2011 provided by the Tokyo Metropolitan Government. The created distribution has a resolution of 5 m, which is sufficiently high to represent streets between buildings (Fig. 2). The initial and boundary conditions for the mesoscale simulations were obtained from grid-point-value data [52] of the Meso-Scale Model at the Japan Meteorological Agency, whereas those for the micrometeorological simulations were obtained from the results of Domain 3.



Figure 2: Computational domains for the mesoscale and micrometeorological simulations. The latter resolves the buildings around Tokyo Station, as shown in the rightmost building height distribution. The satellite images were obtained from MapBox [53] and Open-StreetMap [54]. In the building height distribution, the upper black square, labeled "A", represents the region shown in Fig. 8, and the lower square, labeled "B", represents the region shown in Figs. 4, 5, and 10.

The numerical experiments were conducted for hot days between 2013 and 2020. The focus on hot days was motivated by the possibility that the risk of heat-related illness could be influenced by temperature on street scales [55].

Specifically, 171 hot summer hours were selected between 2013 and 2020 in which the maximum daily temperature exceeded 35°C. Figure 3 shows an instantaneous snapshot of the simulated three-dimensional temperature, which appears to be advected by the flows passing around the buildings. Each simulation was run for each targeted hour. The result from the first 10 minutes was discarded, and the remaining 50 minutes were used to obtain 1-minute average values. These average fields were sampled every 2 minutes. This sampling interval was determined by the decrease in auto-correlation below a threshold of 0.5. Consequently, 25 snapshots in three dimensions were obtained from each experiment, totaling 4,275 snapshots (=  $25 \times 171$ ).



Figure 3: A three-dimensional distribution of the 1-minute average temperature at 2013-07-09 14:00+09:00 from the micrometeorological simulation. The Volume Data Visualizer for Google Earth (VDVGE) [56] was utilized for the visualization.

# 3.2. Data preparation for deep learning

Pairs of HR and LR data were generated for supervised learning. First, we cropped the HR results from the micrometeorological simulations as follows.

The central  $320 \times 320$  grid points were extracted to eliminate the influence of the lateral damping layer for dynamical downscaling; then, vertically, the bottom 32 grid points were selected to focus on atmospheric flows near the ground. As a result, each HR snapshot consists of  $320 \times 320 \times 32$  voxels. Since the resolution is 5 m in all directions, the east-west and north-south widths are 1,600 m and the vertical width is 160 m.

The LR data were obtained by applying the average pooling to the HR data such that the resolution is four times lower, i.e., the 20 m resolution in all directions. The factor of four is sufficiently coarse for practical applications [7, 26, 30]. The LR building height distribution was obtained from the HR distribution. Specifically, the entire  $320 \times 320$  region was divided into  $4 \times 4$ blocks, and each block was replaced with the mean value. This operation is called the average pooling with a  $4 \times 4$  kernel. The total volume of buildings is preserved between the HR and LR, even though some streets between buildings are not represented at the LR, as shown in Fig. 4. This fact follows that the total volume of the atmosphere is preserved as well. The LR temperature and velocity were then created in two steps. First, the average pooling with a  $4 \times 4 \times 4$  kernel was applied to each HR snapshot  $(T^{\text{HR}})$ ,  $u^{\text{HR}}$ ,  $v^{\text{HR}}$ , or  $w^{\text{HR}}$ ). In this average operation, the grid points within the LR buildings were excluded. Second, NaN values were inserted at the grid points of the LR buildings. In the subsequent preprocessing, these NaN values were replaced with zeros (Section 2). The resultant LR snapshots consist of  $80 \times 80 \times 8$  voxels and are four times coarser than the HR snapshots.



Figure 4: Building height distributions over a cropped region: (a) 5 m and (b) 20 m resolution. This region is labeled "B" in Fig. 2.

All pairs of HR and LR data were split into training, validation, and test datasets in proportions of 60%, 20%, and 20%, respectively. This splitting preserved the chronological order to prevent the so-called data leakage. In more detail, the training dataset was obtained from 99 micrometeorological simulations between 2013 and 2018 (2,475 pairs of HR and LR snapshots); the validation set was from 36 simulations between 2018 and 2019 (900 pairs); and the test set was from 36 simulations between 2019 and 2020 (900 pairs).

### 3.3. Training of the CNN

The CNN was trained via supervised learning using the following loss function:

$$\|\boldsymbol{Y}^{\mathbf{HR}} - \hat{\boldsymbol{Y}}^{\mathbf{HR}}\|^{2} + \lambda_{\text{grd}} \|B^{\text{HR}} \odot \nabla (\boldsymbol{Y}^{\mathbf{HR}} - \hat{\boldsymbol{Y}}^{\mathbf{HR}})\|^{2} + \lambda_{\text{div}} \|B^{\text{HR}} \odot (\nabla \cdot \boldsymbol{V}^{\mathbf{HR}} - \nabla \cdot \hat{\boldsymbol{V}}^{\mathbf{HR}})\|^{2}, \qquad (3)$$

where  $\mathbf{Y}^{\mathbf{HR}}$  represents the ground truth,  $\mathbf{Y}^{\mathbf{HR}} = (T^{\mathbf{HR}}, u^{\mathbf{HR}}, v^{\mathbf{HR}}, w^{\mathbf{HR}}) = (T^{\mathbf{HR}}, \mathbf{V}^{\mathbf{HR}})$ ;  $\hat{\mathbf{Y}}^{\mathbf{HR}}$  is the corresponding inference by the CNN;  $\lambda_{\text{grd}}$  and  $\lambda_{\text{div}}$  are positive real constants;  $\|\cdot\|^2$  stands for mean squared quantities;  $\odot$  is element-wise multiplication; and  $\nabla$  denotes the second-order centered difference. The first term in (3) is the mean squared error. The second term is the mean gradient error [39, 57]. The third term is the mean divergence

error [23, 24, 39]. The building mask  $B^{\rm HR}$  is included in the second and third terms to exclude the invalidity of finite differences at the grid points near building boundaries. The first term assesses the flow-field values, whereas the second and third terms measure the difference in the flow-field smoothness. Previous studies reported the effectiveness of the second or third term [23, 24, 39, 57]. In particular, the third term, i.e., the error in the divergence of three-dimensional velocity, can enhance the physical validity of inference. Due to this term, the trained CNN is regarded as a physics-guided NN.

The CNN was trained by using the Adam optimizer [58] with a learning rate of  $1.0 \times 10^{-4}$ . Here, the parameters in the loss function (3) were set at  $\lambda_{\rm grd} = 1$  and  $\lambda_{\rm div} = 10$ . We stored the weights of the CNN that resulted in the smallest validation loss after 1,200 epochs. During training,  $64 \times 64 \times 32$ voxels were randomly cropped from the HR data, and the corresponding region was extracted from the LR data. Repeating this sampling, a minibatch of 32 samples was created and fed to the CNN.

The CNN was implemented with PyTorch 1.11.0 [59] and trained with distributed data parallel (DDP) using two GPU boards of NVIDIA A100 40GB PCIe on the Earth Simulator supercomputer system in Japan Agency for Marine-Earth Science and Technology. During the evaluation, one GPU was used. The CNN implementation is available on the Zenodo repository (see Data availability).

#### 3.4. Evaluation metrics for the CNN

The CNN performance was evaluated using two types of metrics for voxelwise accuracy and pattern consistency, where temperature and velocity were analyzed separately. At the evaluation phase, the inference was obtained by feeding LR snapshots of the full size  $80 \times 80 \times 8$  into the trained CNN.

The mean voxel-wise errors in temperature and velocity are defined as

follows:

$$|\Delta T| = \frac{1}{N} \sum_{\text{const height}} B^{\text{HR}} \left| T^{\text{HR}} - \hat{T}^{\text{HR}} \right|,$$
$$= \frac{1}{N} \sum_{\text{const height}} B^{\text{HR}} \sqrt{\left( T^{\text{HR}} - \hat{T}^{\text{HR}} \right)^2}, \tag{4}$$

and

$$|\Delta \boldsymbol{V}| = \frac{1}{N} \sum_{\text{const height}} B^{\text{HR}} \left| \boldsymbol{V}^{\text{HR}} - \hat{\boldsymbol{V}}^{\text{HR}} \right|,$$
  
$$= \frac{1}{N} \sum_{\text{const height}} B^{\text{HR}} \sqrt{(u^{\text{HR}} - \hat{u}^{\text{HR}})^2 + (v^{\text{HR}} - \hat{v}^{\text{HR}})^2 + (w^{\text{HR}} - \hat{w}^{\text{HR}})^2},$$
(5)

where the summation is taken at a constant height over the horizontal sections in all test data, and N denotes the number of the grid points outside buildings, i.e.,  $N = \sum_{\text{const height}} B^{\text{HR}}$ . Equations (4) and (5) are referred to as the error norm of temperature and velocity, respectively.

The pattern consistency is estimated with the mean structural similarity index measure (MSSIM) [60]:

MSSIM loss = 1 - MSSIM,  
= 
$$1 - \frac{1}{N} \sum_{\text{const height}} \left[ \frac{(2\mu\hat{\mu} + C_1)(2\chi + C_2)}{(\mu^2 + \hat{\mu}^2 + C_1)(\sigma^2 + \hat{\sigma}^2 + C_2)} \right],$$
 (6)

where  $C_1 = 0.01$ ,  $C_2 = 0.03$ ,  $\mu$  and  $\sigma^2$  are, respectively, the mean and variance of the ground truth;  $\hat{\mu}$  and  $\hat{\sigma}^2$  are the corresponding quantities of the inference; and  $\chi$  is the covariance between the ground truth and inference. The variables in the summation are calculated locally by applying a three-dimensional Gaussian filter along with the building mask  $B^{\text{HR}}$ . The MSSIM loss takes a value larger than or equal to 0, and smaller values indicate that

the spatial patterns of inference are more similar to those of the ground truth. A detailed discussion on MSSIM is found in Wang et al. [60]. Equation (6) was applied directly to temperature fields. For velocity, (6) was applied to each component of u, v, and w, and the obtained three MSSIM losses were averaged. We confirmed that the difference in the MSSIM loss among the velocity components is sufficiently small ( $\leq 0.05$ ).

# 4. Results and discussion

# 4.1. Evaluation of the CNN inference

We show that the CNN successfully reconstructed HR air temperature and wind velocity fields. Figure 5 shows an example of the inference at the 3 and 43 m heights. The ground elevation is not constant and its variation is incorporated in the micrometeorological simulations. For simplicity, however, this variation is ignored in the analysis because it is less than 5 m, and the height measured from the ground is used. The gray in the figure represents the missing regions within the buildings. At the 3 m height, most streets between the buildings are not represented in the LR input (Fig. 5a). Nevertheless, the CNN successfully inferred the HR temperature and velocity fields, demonstrating its ability to simultaneously perform the enhancement of spatial resolution and the reconstruction of missing data. At the 43 m height (Fig. 5b), there are fewer grid points of buildings, and the CNN shows an inference similar to the ground truth. Ordinary SR by CNNs can be interpreted as pattern matching between HR and LR images [10]. Our result implies that this pattern matching allows CNNs to super-resolve three-dimensional flow fields while restoring missing data.



Figure 5: An example of the inference at (a) 3 m and (b) 43 m heights: air temperature (T) and the eastward, northward, and upward components of wind velocity (u, v, and w, respectively). The building grids are colored gray in the figure. In (b), the building edges are enclosed by black lines to emphasize the presence of a few buildings. The region shown here is labeled "B" in Fig. 2.

The quantitative analysis confirmed the high accuracy of the CNN inference. Figure 6 shows the test metrics at all heights. As the height decreases (i.e., closer to the ground), the error norm and MSSIM loss increase, likely due to the increase in the number of building grid points. For comparison, we employed a baseline model that uses linear interpolation and extrapolation. It is obvious that this baseline model failed to infer the temperature and velocity near the ground. Indeed, the test metrics of the baseline were approximately one magnitude larger than those of the CNN (Fig. 6). This result suggests that traditional methods such as linear interpolation are not capable of performing three-dimensional SR while restoring missing data. As the height increases, the inference accuracy improves for both the linear model and CNN, likely due to the decrease in the number of building grid points. Even at upper heights, the CNN was more accurate than the linear model. The test metrics of the CNN take the following maxima at the lowest (3 m) height level (Fig. 6): for temperature T, the error norm is about 0.10 K and the MSSIM loss is about 0.06; for velocity V, the error norm is about 0.35 m s<sup>-1</sup> and the MSSIM loss is about 0.07. Over all heights, the MSSIM loss for T and V is smaller than 0.1. Such small MSSIM loss values suggest the similarity between two patterns [60]. Thus, the CNN reconstructed the three-dimensional flow patterns over the entire height range.



Figure 6: Dependence of the test metrics on height. The error norms and MSSIM losses are defined in (4), (5), and (6). For comparison, we employed a baseline model that uses linear interpolation and extrapolation, which is labeled "Linear."

Figure 7 shows histograms of voxel-wise errors at the 3 and 43 m heights. These errors are the quantities in the summations of (4) and (5). Specifically, the temperature error is given by the absolute value of the temperature difference, i.e.,  $|T^{\text{HR}} - \hat{T}^{\text{HR}}|$  in (4), which has peaks around zeros in Fig. 7. The velocity error is equal to the Euclidean norm of the difference in velocity

vectors, i.e.,  $|\mathbf{V}^{\mathbf{HR}} - \hat{\mathbf{V}}^{\mathbf{HR}}|$  in (5), which has peaks slightly away from zeros in Fig. 7 because the error consists of the sum of squared quantities. The 3 m height is closest to the ground, resulting in wider error distributions compared to 43 m. At the 3 m height, the 90-percentile errors are 0.23 K for T and 0.64 m s<sup>-1</sup> for  $\mathbf{V}$ , while at the 43 m height, 0.07 K for T and 0.46 m s<sup>-1</sup> for  $\mathbf{V}$ . The scaling parameters in (2) are 8.4 K for T and 14.3 m s<sup>-1</sup> for  $\mathbf{V}$ . The latter is the average of the parameters for u, v, and w. Even at the 3 m height, the 90-percentile errors are 2.7% and 4.5% of these scales for temperature and velocity, respectively. Thus, the voxel-wise error in the CNN inference is sufficiently small.



Figure 7: Histograms of the voxel-wise errors in temperature and velocity at (a) 3 m and (b) 43 m heights. The temperature error is  $|T^{\text{HR}} - \hat{T}^{\text{HR}}|$  in (4) and the velocity error is  $|V^{\text{HR}} - \hat{V}^{\text{HR}}|$  in (5). In each histogram, 100,000 error values were randomly sampled from the non-missing grid points in all test results. The dashed lines represent the 90 percentile values.

The accuracy of temperature inference near the ground is comparable to that of our previous study [30]. The present error norm was 0.10 K at the 3 m height (Fig. 6), while the root-mean-square error (RMSE) was 0.15 K for the 2 m height in our previous study [30]. In that study, twodimensional temperatures were generated by sampling grid points at 2 m above the bottom surface (e.g., the ground or roofs of buildings). In such a manner of creating two-dimensional data, no missing value appeared. For direct comparison, two-dimensional temperature fields were obtained in the same manner from both the ground truth and CNN inference. Figure 8 shows an example of the resulting temperature. The region shown in the figure is located at Tokyo Station and is characterized by multiple railway platforms, as seen in the building height distribution. The temperature is lower at points of taller buildings because the temperatures above these buildings were sampled, leading to the discontinuities in temperature (Fig. 8). The CNN reproduced these discontinuities, which indicates the accuracy of the base, three-dimensional inference. Furthermore, the mean absolute error of the two-dimensional temperature is 0.13 K, whereas the corresponding error was 0.11 K in our previous study [30]. The proposed CNN exhibits comparable accuracy, even though it is not specifically designed to infer the two-dimensional temperature.



Figure 8: An example of the two-dimensional temperature field at the 2 m height above the bottom surface (e.g., the ground or roofs of buildings). The region shown here is located at Tokyo Station, where multiple railway platforms are present. This region is labeled "A" in Fig. 2. Our previous study shows an inference at the same location (Fig. 6 in [30]).

The computation time for the CNN inference is negligibly small, implying that the CNN can be employed as an NN in SR simulation systems [7, 25, 26]. For instance, it took approximately 0.27 seconds to generate an inference on the full size of  $320 \times 320 \times 32$  voxels. The wall time for SR simulations can be estimated as follows. An LR micrometeorological simulation for 1-hour predictions would be conducted, which generates 60 sets of 1-minute average

data. The elapsed time of this simulation was estimated at 312 seconds [30]. The CNN would then need 16.2 (=  $0.27 \times 60$ ) seconds to super-resolve all LR results. Thus, the total wall time for the SR simulation would be about 328 seconds. This estimated time is much shorter than the elapsed time for the HR micrometeorological simulation, which is approximately 8 hours, highlighting the computational efficiency of SR simulation.

## 4.2. Inference only utilizing upper flows

We demonstrate that near-surface flows can be inferred from upper flows. As previously discussed, the simultaneous enhancement of spatial resolution and restoration of missing regions can be interpreted as pattern matching between LR and HR flows. In the vicinity of the ground, the LR input contains many missing grid points due to the presence of buildings (Fig. 5a). In contrast, at upper levels, such missing points become fewer (Fig. 5b). This result suggests that flows at upper levels are more informative for pattern matching. In addition, the spatial scale of upper flows is larger due to the absence of buildings, compared to near-surface flows. This implies that the small-scale flows near the ground can be inferred from the large-scale upper flows. To examine this hypothesis, we made all grid points missing below a certain height in the LR input and conducted the training and evaluation of the CNN.

We first determined an appropriate height threshold for the replacement with missing values. Figure 9 shows a histogram of the building height. The peak of the histogram is around 30 m, which is due to the height cap of 31 m in the old Building Standards Law [61]. The figure indicates that most buildings are lower than 43 m in height. Thus, we replaced all values below 43 m in the LR input with the missing values, i.e., zeros.



Figure 9: Histogram of the building height around Tokyo Station. The corresponding building-height spatial distribution is shown in Fig. 2.

The CNN can qualitatively infer near-surface flows from upper flows, as demonstrated in Fig. 10. The LR input is not shown at the 3 m height (Fig. 10a) since all values are missing. Despite these missing data, the CNN reproduced the flow structures that were qualitatively similar to those of the ground truth. In particular, the CNN accurately identified the locations of small or large values of temperature and velocity, although the detailed patterns differed from those of the ground truth. At the 43 m height (Fig. 10b), no replacement with the missing values was performed in the LR input, which is likely to lead to a more accurate inference, compared to Fig. 10a.



Figure 10: An example of the CNN inference at (a) 3 m and (b) 43 m heights: air temperature (T) and the eastward, northward, and upward components of wind velocity (u, v, and w, respectively). All input values below the 43 m height were made missing and the CNN was then trained and evaluated. The LR input is not shown in (a), as all values are missing. The building grids are colored gray in the figure. In (b), the building edges are enclosed by black lines to emphasize the presence of a few buildings. The region shown here is the same as in Fig. 5 and is labeled "B" in Fig. 2.

Figure 11 compares the accuracy of inference with and without the missingvalue replacement. The case without the replacement has been examined in Fig. 6. The vertical dashed lines in Fig. 11 represent the 43 m height. Above this threshold, the test metrics are comparable for temperature and velocity with and without the missing-value replacement. However, at lower heights, the error norm and MSSIM loss are larger in the case of the missing-value replacement. In particular, at the 3.0 m height (i.e., the bottom-most level) this replacement increased the error norm of temperature from 0.10 to 0.17 K and the MSSIM loss from 0.06 to 0.09, whereas for velocity, the error norm increased from 0.35 to 0.59 m s<sup>-1</sup> and the MSSIM loss from 0.07 to 0.13. Although the error norm and MSSIM loss are approximately 1.5 to 2.0 times larger, the MSSIM loss still remains close to 0.1, leading to the MSSIM being around 0.9 according to (6). When the MSSIM is greater than about 0.9, two images are perceptually similar [60]. Therefore, we conclude that near-surface flows can be qualitatively inferred by the CNN from flows at upper heights.



Figure 11: Dependence of the test metrics on height. The error norms and MSSIM losses are defined in (4), (5), and (6). The CNN was trained and evaluated using the LR input where all values below the 43 m height were made missing. This height threshold is denoted by the dashed vertical lines. The obtained result is labeled "Missing below 43.0 m." The result of the original CNN is denoted by "No replacement," whose error values are the same as in Fig. 6.

#### 5. Conclusions

This study has proposed a convolutional neural network (CNN) that super-resolves three-dimensional air temperature and wind velocity fields obtained from building-resolving micrometeorological simulations. The CNN architecture is based on a U-Net for image inpainting [39]. The superresolution (SR) process in the CNN requires not only increasing the spatial resolution but also restoring the missing regions caused by buildings because the building shape is different between low and high resolution (LR and HR). The proposed CNN incorporates an image inpainting technique, called gated convolution [43]. Additionally, by clipping the input to the range of [0, 1], zeros are regarded as missing values. This preprocessing eliminates the need for LR building information and allows the CNN to perform inference using only the HR building data.

The CNN performance has been evaluated via supervised learning using the results of micrometeorological simulations for an actual urban area around Tokyo Station in Japan. The LR input data were created by applying the average pooling to the HR simulation results. The training loss included a physics loss term, namely the error in the divergence of velocity vectors. The CNN reconstructed the velocity and temperature fields with high accuracy. The error was larger near the ground than at upper heights, which is likely due to the missing data in the input caused by the coarseness of the LR buildings. The mean errors at the lowest height were approximately 0.10 K for temperature and  $0.35 \text{ m s}^{-1}$  for velocity. Furthermore, we have demonstrated that small-scale near-surface flows can be qualitatively inferred from the large-scale flows above most buildings. This result implies that flows around buildings can be estimated from the LR simulation results in which buildings are not accurately represented.

There are at least two potential directions for future research. The first direction is to investigate the generalizability of the proposed CNN to other cities. Our previous study [30] demonstrated that the SR model for the twodimensional temperature can be applied to a city different from the city used in training. It would be interesting to examine whether similar results are obtained with the proposed CNN for the three-dimensional SR. The second direction is to explore the generation of LR and HR data from separate simulations, since in the concepts of SR simulation [7, 25, 26], LR and HR data are obtained from LR and HR atmospheric simulations, respectively, and a neural network is trained using these data. The effectiveness of SR simulation has been shown on meteorological or climate scale [25, 26]. In urban micrometeorology, however, it is necessary to incorporate changes in building shapes during the SR process. Based on the proof of concept in this paper, it may be possible to develop an SR simulation system for real-time micrometeorological predictions in urban cities.

# Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

The source code for deep learning is preserved at the Zenodo repository (https://doi.org/10.5281/zenodo.7777210) and developed openly at the GitHub repository (https://github.com/YukiYasuda2718/3d-sr-micrometeorology). The data that support the findings of this study are available from the corresponding author upon reasonable request.

# Acknowledgements

This work was supported by the JSPS KAKENHI (grant number 20H05751). The micrometeorological simulation and deep learning were performed on the Earth Simulator system (project ID: 0-22021) at the Japan Agency for Marine-Earth Science and Technology (JAMSTEC).

# References

- U. N. H. S. Programme, World Cities Report 2020, United Nations, 2020. doi:10.18356/27bc31a5-en. URL https://www.un-ilibrary.org/content/books/ 9789210054386
- [2] D. Mauree, E. Naboni, S. Coccolo, A. Perera, V. M. Nik, J.-L. Scartezzini, A review of assessment methods for the urban environment and its energy sustainability to guarantee climate adaptation of future cities, Renewable and Sustainable Energy Reviews 112 (2019) 733-746. doi:https://doi.org/10.1016/j.rser.2019.06.005. URL https://www.sciencedirect.com/science/article/pii/S1364032119303909
- [3] D. Lai, W. Liu, T. Gan, K. Liu, Q. Chen, A review of mitigating strategies to improve the thermal environment and thermal comfort in urban outdoor spaces, Science of The Total Environment 661 (2019) 337-353. doi:https://doi.org/10.1016/j.scitotenv.2019.01.062.
   URL https://www.sciencedirect.com/science/article/pii/ S0048969719300683
- [4] E. Naboni, Environmental simulation tools in architectural practice.: the impact on processes, methods and design., in: Sustainable architecture for a renewable future, Fraunhofer IRB Verlag, 2013, pLEA 2013; Conference date: 10-09-2013 Through 12-09-2013.
- [5] Y. Toparlar, B. Blocken, B. Maiheu, G. J. van Heijst, A review on the cfd analysis of urban microclimate, Renewable and Sustainable Energy Reviews 80 (2017) 1613–1640. doi:10.1016/j.rser.2017.05.248.
- [6] C. K. C. Lam, H. Lee, S.-R. Yang, S. Park, A review on the significance and perspective of the numerical simulations of outdoor

thermal environment, Sustainable Cities and Society 71 (2021) 102971. doi:https://doi.org/10.1016/j.scs.2021.102971.

URL https://www.sciencedirect.com/science/article/pii/ S2210670721002572

- [7] R. Onishi, D. Sugiyama, K. Matsuda, Super-resolution simulation for real-time prediction of urban micrometeorology, SOLA 15 (2019) 178– 182. doi:10.2151/sola.2019-032.
- [8] D. Kodi Ramanah, T. Charnock, F. Villaescusa-Navarro, B. D. Wandelt, Super-resolution emulator of cosmological simulations using deep physical models, Monthly Notices of the Royal Astronomical Society 495 (4) (2020) 4227-4236. arXiv:https://academic.oup.com/mnras/article-pdf/495/4/4227/33372045/staa1428.pdf, doi: 10.1093/mnras/staa1428. URL https://doi.org/10.1093/mnras/staa1428
- [9] K. Shiina, H. Mori, Y. Tomita, H. K. Lee, Y. Okabe, Inverse renormalization group based on image super-resolution using deep convolutional networks, Scientific Reports 11 (12 2021). doi:10.1038/ s41598-021-88605-w.
- [10] C. Dong, C. C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), Computer Vision – ECCV 2014, Springer International Publishing, Cham, 2014, pp. 184–199.
- [11] V. K. Ha, J.-C. Ren, X.-Y. Xu, S. Zhao, G. Xie, V. Masero, A. Hussain, Deep learning based single image super-resolution: A survey, International Journal of Automation and Computing (2019). doi: 10.1007/s11633-019-1183-x.
- [12] S. Anwar, S. Khan, N. Barnes, A deep journey into super-resolution: A

survey, ACM Comput. Surv. 53 (3) (may 2020). doi:10.1145/3390462. URL https://doi.org/10.1145/3390462

- [13] A. Ducournau, R. Fablet, Deep learning for ocean remote sensing: an application of convolutional neural networks for super-resolution on satellite-derived sst data, 2016, pp. 1–6. doi:10.1109/PRRS.2016. 7867019.
- [14] Y. Xie, E. Franz, M. Chu, N. Thuerey, Tempogan: A temporally coherent, volumetric gan for super-resolution fluid flow, ACM Trans. Graph. 37 (7 2018). doi:10.1145/3197517.3201304. URL https://doi.org/10.1145/3197517.3201304
- [15] Z. Deng, C. He, Y. Liu, K. C. Kim, Super-resolution reconstruction of turbulent velocity fields using a generative adversarial network-based artificial intelligence framework, Physics of Fluids 31 (12 2019). doi: 10.1063/1.5127031.
- [16] K. Fukami, K. Fukagata, K. Taira, Super-resolution reconstruction of turbulent flows with machine learning, Journal of Fluid Mechanics 870 (2019) 106–120. doi:10.1017/jfm.2019.238.
- [17] K. Fukami, K. Fukagata, K. Taira, Machine-learning-based spatiotemporal super resolution reconstruction of turbulent flows, Journal of Fluid Mechanics 909 (2021) A9. doi:10.1017/jfm.2020.948.
- [18] L. Wang, Z. Luo, J. Xu, W. Luo, J. Yuan, A novel framework for cost-effectively reconstructing the global flow field by super-resolution, Physics of Fluids 33 (9) (2021) 095105. arXiv:https://doi.org/10. 1063/5.0062775, doi:10.1063/5.0062775. URL https://doi.org/10.1063/5.0062775
- [19] O. Obiols-Sales, A. Vishnu, N. P. Malaya, A. Chandramowlishwaran, Surfnet: Super-resolution of turbulent flows with transfer learning us-

ing small datasets, in: 2021 30th International Conference on Parallel Architectures and Compilation Techniques (PACT), IEEE Computer Society, Los Alamitos, CA, USA, 2021, pp. 331-344. doi: 10.1109/PACT52795.2021.00031. URL https://doi.ieeecomputersociety.org/10.1109/PACT52795.

- 2021.00031
- [20] O. Obiols-Sales, A. Vishnu, N. Malaya, A. Chandramowlishwaran, Nunet: Deep learning for non-uniform super-resolution of turbulent flows (2022). doi:10.48550/ARXIV.2203.14154.
   URL https://arxiv.org/abs/2203.14154
- [21] C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, K. Kashinath, M. Mustafa, H. A. Tchelepi, P. Marcus, M. Prabhat, A. Anandkumar, Meshfreeflownet: A physics-constrained deep continuous space-time super-resolution framework, in: 2020 SC20: International Conference for High Performance Computing, Networking, Storage and Analysis (SC), IEEE Computer Society, Los Alamitos, CA, USA, 2020, pp. 1–15. doi:10.1109/SC41405.2020.00013. URL https://doi.ieeecomputersociety.org/10.1109/SC41405.

2020.00013

- [22] C. Wang, E. Bentivegna, W. Zhou, L. Klein, B. Elmegreen, Physicsinformed neural network super resolution for advection-diffusion models, in: Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020), 2020. URL https://ml4physicalsciences.github.io/2020/
- [23] M. Bode, M. Gauding, Z. Lian, D. Denker, M. Davidovic, K. Kleinheinz, J. Jitsev, H. Pitsch, Using physics-informed enhanced super-resolution generative adversarial networks for subfilter modeling in turbulent reactive flows, Proceedings of the Combustion Institute 38 (2021) 2617–2625. doi:10.1016/J.PROCI.2020.06.022.

[24] T. Bao, S. Chen, T. T. Johnson, P. Givi, S. Sammak, X. Jia, Physics guided neural networks for spatio-temporal super-resolution of turbulent flows, in: The 38th Conference on Uncertainty in Artificial Intelligence, 2022.

URL https://openreview.net/forum?id=S98VJL8jcxq

- [25] Y. Wu, B. Teufel, L. Sushama, S. Belair, L. Sun, Deep learning-based super-resolution climate simulator-emulator framework for urban heat studies, Geophysical Research Letters 48 (19) (2021) e2021GL094737, e2021GL094737 2021GL094737. arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2021GL094737, doi:https://doi.org/10.1029/2021GL094737.
  URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GL094737
- [26] J. Wang, Z. Liu, I. Foster, W. Chang, R. Kettimuthu, V. R. Kotamarthi, Fast and accurate learned multiresolution dynamical downscaling for precipitation, Geoscientific Model Development 14 (10) (2021) 6355– 6372. doi:10.5194/gmd-14-6355-2021. URL https://gmd.copernicus.org/articles/14/6355/2021/
- [27] T. Vandal, E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, A. R. Ganguly, Deepsd: Generating high resolution climate change projections through single image super-resolution, Association for Computing Machinery, 2017, pp. 1663–1672. doi:10.1145/3097983.3098004. URL https://doi.org/10.1145/3097983.3098004
- [28] E. R. Rodrigues, I. Oliveira, R. Cunha, M. Netto, Deepdownscale: A deep learning strategy for high-resolution weather forecast, 2018, pp. 415-422. doi:10.1109/eScience.2018.00130.
- [29] K. Stengel, A. Glaws, D. Hettinger, R. N. King, Adversarial superresolution of climatological wind and solar data, Proceedings of the Na-

tional Academy of Sciences 117 (2020) 16805-16815. doi:10.1073/ PNAS.1918964117. URL https://www.pnas.org/content/117/29/16805

- [30] Y. Yasuda, R. Onishi, Y. Hirokawa, D. Kolomenskiy, D. Sugiyama, Super-resolution of near-surface temperature utilizing physical quantities for real-time prediction of urban micrometeorology, Building and Environment 209 (2022) 108597. doi:https: //doi.org/10.1016/j.buildenv.2021.108597.
  URL https://www.sciencedirect.com/science/article/pii/ S0360132321009884
- [31] K. Fukami, K. Fukagata, K. Taira, Super-resolution analysis via machine learning: A survey for fluid flows (2023). doi:10.48550/ARXIV.2301. 10937. URL https://arxiv.org/abs/2301.10937
- [32] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, Y. Akbari, Image inpainting: A review, Neural Processing Letters 51 (2020) 2007–2028. doi:10.1007/s11063-019-10163-0.
- [33] Z. Qin, Q. Zeng, Y. Zong, F. Xu, Image inpainting based on deep learning: A review, Displays 69 (2021) 102028. doi:https://doi.org/10.1016/j.displa.2021.102028. URL https://www.sciencedirect.com/science/article/pii/ S0141938221000391
- [34] W. Li, Y. Li, D. Chen, J. C.-W. Chan, Thin cloud removal with residual symmetrical concatenation network, ISPRS Journal of Photogrammetry and Remote Sensing 153 (2019) 137-150. doi:https://doi.org/10.1016/j.isprsjprs.2019.05.003.
  URL https://www.sciencedirect.com/science/article/pii/S092427161930125X

- [35] C. Kadow, D. M. Hall, U. Ulbrich, Artificial intelligence reconstructs missing climate information, Nature Geoscience 13 (2020) 408–413. doi: 10.1038/s41561-020-0582-5.
- [36] A. Meraner, P. Ebel, X. X. Zhu, M. Schmitt, Cloud removal in sentinel-2 imagery using a deep residual neural network and sar-optical data fusion, ISPRS Journal of Photogrammetry and Remote Sensing 166 (2020) 333– 346. doi:https://doi.org/10.1016/j.isprsjprs.2020.05.013. URL https://www.sciencedirect.com/science/article/pii/ S0924271620301398
- [37] A. Geiss, J. C. Hardin, Inpainting radar missing data regions with deep learning, Atmospheric Measurement Techniques 14 (12) (2021) 7729– 7747. doi:10.5194/amt-14-7729-2021. URL https://amt.copernicus.org/articles/14/7729/2021/
- [38] D. Maduskar, N. Dube, Navier-stokes-based image inpainting for restoration of missing data due to clouds, in: M. K. Sharma, V. S. Dhaka, T. Perumal, N. Dey, J. M. R. S. Tavares (Eds.), Innovations in Computational Intelligence and Computer Vision, Springer Singapore, Singapore, 2021, pp. 497–505.
- [39] L. Schweri, S. Foucher, J. Tang, V. C. Azevedo, T. Günther, B. Solenthaler, A physics-aware neural network approach for flow data reconstruction from satellite observations, Frontiers in Climate 3 (2021). doi:10.3389/fclim.2021.656505.
  URL https://www.frontiersin.org/articles/10.3389/fclim. 2021.656505
- [40] M. Czerkawski, P. Upadhyay, C. Davison, A. Werkmeister, J. Cardona, R. Atkinson, C. Michie, I. Andonovic, M. Macdonald, C. Tachtatzis, Deep internal learning for inpainting of cloud-affected regions in satellite

imagery, Remote Sensing 14 (6) (2022). doi:10.3390/rs14061342. URL https://www.mdpi.com/2072-4292/14/6/1342

- [41] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham, 2015, pp. 234–241.
- [42] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, B. Catanzaro, Image inpainting for irregular holes using partial convolutions, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018.
- [43] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, T. S. Huang, Free-form image inpainting with gated convolution, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
- [44] A. L. Maas, A. Y. Hannun, A. Y. Ng, et al., Rectifier nonlinearities improve neural network acoustic models, in: Proc. icml, Vol. 30, Atlanta, Georgia, USA, 2013, p. 3.
- [45] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [46] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Los Alamitos, CA, USA, 2016, pp. 1874–1883. doi:10.1109/CVPR.2016.207. URL https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.

207

- [47] K. Takahashi, R. Onishi, Y. Baba, S. Kida, K. Matsuda, K. Goto, H. Fuchigami, Challenge toward the prediction of typhoon behaviour and down pour, Journal of Physics: Conference Series 454 (2013) 012072. doi:10.1088/1742-6596/454/1/012072. URL https://iopscience.iop.org/article/10.1088/1742-6596/ 454/1/012072
- [48] R. Onishi, K. Takahashi, A warm-bin-cold-bulk hybrid cloud microphysical model\*, Journal of the Atmospheric Sciences 69 (2012) 1474-1497. doi:10.1175/JAS-D-11-0166.1.
   URL https://journals.ametsoc.org/doi/10.1175/ JAS-D-11-0166.1
- [49] W. Sasaki, R. Onishi, H. Fuchigami, K. Goto, S. Nishikawa, Y. Ishikawa, K. Takahashi, Mjo simulation in a cloud-system-resolving global ocean-atmosphere coupled model, Geophysical Research Letters 43 (2016) 9352-9360. doi:https://doi.org/10.1002/2016GL070550.
  URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL070550
- [50] K. Matsuda, R. Onishi, K. Takahashi, Tree-crown-resolving large-eddy simulation coupled with three-dimensional radiative transfer model, Journal of Wind Engineering and Industrial Aerodynamics 173 (2018) 53-66. doi:10.1016/j.jweia.2017.11.015.
- J. Gutowski, [51] F. Giorgi, W. Regional dynamical downscaling and the cordex initiative, Annual Review of Environment and Resources 40 (1) (2015) 467-490. arXiv:https: //doi.org/10.1146/annurev-environ-102014-021217, doi: 10.1146/annurev-environ-102014-021217. URL https://doi.org/10.1146/annurev-environ-102014-021217

- [52] Japan meteorological business support center. URL http://www.jmbsc.or.jp/en/index-e.html
- [53] Mapbox. URL https://www.mapbox.com/about/maps/
- [54] Openstreetmap. URL https://www.openstreetmap.org/about/
- [55] T. Kamiya, R. Onishi, S. Kodera, A. Hirata, Estimation of timecourse core temperature and water loss in realistic adult and child models with urban micrometeorology prediction, International Journal of Environmental Research and Public Health 16 (24) (2019). doi: 10.3390/ijerph16245097. URL https://www.mdpi.com/1660-4601/16/24/5097
- [56] S. Kawahara, R. Onishi, K. Goto, K. Takahashi, Realistic representation of clouds in google earth, in: SIGGRAPH Asia 2015 Visualization in High Performance Computing, SA '15, Association for Computing Machinery, New York, NY, USA, 2015. doi:10.1145/2818517.2818541. URL https://doi.org/10.1145/2818517.2818541
- [57] Z. Lu, Y. Chen, Single image super-resolution based on a modified u-net with mixed gradient loss, Signal, Image and Video Processing 16 (2022) 1143–1151. doi:10.1007/s11760-021-02063-5.
- [58] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2015.
   URL http://arxiv.org/abs/1412.6980
- [59] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner,

L. Fang, J. Bai, S. Chintala, Pytorch: An imperative style, highperformance deep learning library, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems 32, Curran Associates, Inc., 2019, pp. 8024-8035. URL http://papers.neurips.cc/paper/ 9015-pytorch-an-imperative-style-high-performance-deep-learning-library. pdf

- [60] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing 13 (4) (2004) 600–612. doi:10.1109/TIP.2003.819861.
- [61] J. Song, Global Tokyo, Springer, 2020. doi:10.1007/ 978-981-15-3495-9.