Group Project

Ke Ma, Yuhong Li, Fabian Kresse

***Abstract*—Urbanization is the population shift from the rural areas to urban ones. Fast urbanization plays a key role in developing human societies. And the acceleration of urbanization in 21 century has raised researchers interests and concerns in aspect of geography, sociology, public health, education, etc. Since population moving into cities across the globe are exploding in size, new immigrants may not afford high living expense and be forced to settle in the slums, which lack of the spatial accesses and related service such as water, sanitation, emergency services, etc. In this report, we first summarize existing works that utilize heuristic-based algorithms to find the solutions that grow the street networks, re-planning the buildings in the existing slums. Then, we approach the urban planning with a novel deep learning algorithm. By learning the time-lapse of previous city development, the neural network can predict the possible urban planning when assigned the region of interest (ROI) on the satellite maps. We also show the results of our algorithm and demonstrate the potential usage. Our code is available here.**

I. Introduction

The first evidences of urban planning reach as far back as 2500 BCE when pre-modern civilizations engaged in designing and planning large-scale cities [3]. Especially in the last century, in which the population in many parts of the world has grown and still grows at unprecedented rates, the need for well-thought-out urban planning has become ever more important. Furthermore, in recent years there has been an unprecedented influx of data regarding urban development, as evident by large dataset such as Spacenet-7 [8]. Analyzing this data calls for computational methods combined with recent advances in artificial intelligence which offer countless novel opportunities to perform urban planning at an unprecedented level.

One especially important application of urban planning is improving access to public infrastructure, and hence living conditions, in especially poor and underdeveloped neighborhoods like slums. As pointed out by Brelsford et al. almost three billion people could live in slums by 2050 if no countermeasures are taken [7]. As discussed by the authors in [7] there are essentially two approaches to improving and evolving slums: complete reconstruction through *master-planning* and *gradual local evolution*. Due to various reasons, Brelsford et al. suggest an approach of gradual local evolution that enables access to infrastructure at a minimal disruption in the local neighborhood. Hence, the authors of [7] propose a technique that allows optimal reblocking of neighborhoods based on a topological approach.

This project report suggest one extension to their approach. The proposed extension attempts to predict future developments and growths in slums, such that prescient measures can be taken. Some preliminary work regarding this has been performed for this project and is presented in this report. This report is organized as follows: First, section II discusses the work done in [6] and [7], since this work draws some inspiration from it. Section III highlights the proposed extension. Section V discusses some preliminary results regarding our solution to predict urban development, using UNet [4] and an adaption of it. Last, section VI gives a conclusion to this report.

II. Background

The first step the authors of [7] take in order to quantify access to infrastructure is transforming the geometric problem of analyzing city blocks into a topological one utilizing graph theory. This transformation is performed by successively constructing weak dual graphs of the original problem. Hence, the procedure works by representing parcels (i.e., faces of the graph) as nodes and connecting neighboring parcels in the topological graph with edges in the first step. This procedure is performed until there is only one remaining face in the graph (i.e. the graph is a tree). The block complexity k is subsequently yielded by the number of the above steps required. As argued in [7] the block complexity gives a quantitative way to estimate how difficult access to the hardest to access parts of a block are. Furthermore, through systematic reduction of k, it is possible to archive an access network of minimum total length to all infrastructure in a block. Specifically, it seeks the new shortest path to access interior parcels at each step, which can be operated in both strict or statistical method. This procedure is referred as topological optimization. This results in an access network that travel distance between two spots can be much larger than real geometric distance as topological optimization produces tree graphs. Therefore, a $n \times n$ (n is the total numbers of parcels) travel distance matrix $\mathcal{T}$ is derived to find an optimal access network that minimizes average travel distance. Typically, we start from the parcel with a minimum ratio of geometric distance to travel distance and connect it with a nearby node on the access network. Then we update $\mathcal{T}$ and repeated this process until a desired average travel distance is archived. This procedure is the geometric optimization. The approach in [6] is quite similar to [7].

III. Predicting urban development with UNet

As discussed previously, we aim to predict urban development such that measures, such as applying the approaches in [6] and [7] can be taken presciently.

Our approach to do this utilizes UNet - a convolutional neural network model which performs semantic image segmentation [4]. Furthermore, we use the publicly available Spacenet-7 dataset [8]. This dataset contains monthly satellite images of various urban areas on a timescale of two years. In

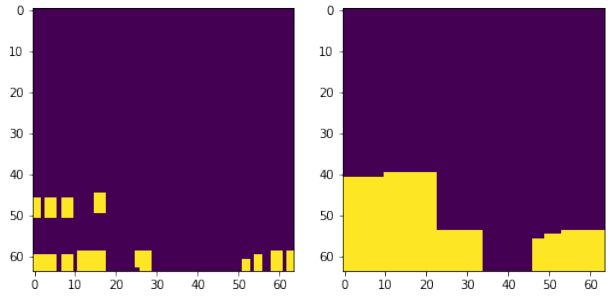

Fig. 1. The left image shows the ground truth of the change, used for calculating the loss of the model. The right image shows the blurred mask supplied as an input to the model.

addition, it provides labeled masks that highlight buildings. In our attempt to predict future urban development, we first used the original UNet and later adapted it in order to show further potential for future work. Working with the original UNet, we decided on a 4 channel input feature map and trained it to output a 1 channel mask predicting future urban development. The input feature map consists of a three channel RGB satellite image and an additionally concatenated mask of likely future building spaces or region of interest (ROI). This leads to one of the current weaknesses of our approach, since this mask is generated from the ground truth label, which would not be available in a real context. We first apply a Gaussian filter to the ground truth change, then we use a threshold to binarize the mask in order to highlight potential areas for the model. A visualization of this can be seen in Figure 1.

While this approach gives the model additional information it would not have in a real setting, we still believe that it learns important features and future adaptions will be able to overcome these limitations. Our reasoning for this is laid out in section V. However, we want to stress that we are aware that the model performs potentially orders of magnitude better than it would without this knowledge of the ground truth.

For our second approach, we modified UNet (henceforth called augmented UNet) and feed it with two four channel input feature maps. The additional feature map shows an image of the same urban area one year previous to the other image. The concatenated mask is the same as in the previous image, such that no biases are introduced this way. Our adapted UNet can be seen in the Appendix (Figure 4).

We measured the quality of the model with the Sørensen-Dice metric [1, 2], which is computed as follows (with TP being the amount of correctly classified pixels, FP the amount of false positive and FN the amount of false negatives):

$$\frac{2 * TP}{2 * TP + FP + FN} \tag{1}$$

We choose this metric since it accounts for false positive and false negatives in addition to true positives. Furthermore, the metric is easy to compute and appears (as judged visually by the authors) to correlate with the quality of the output.

IV. Experiment Settings

We utilize the SpaceNet-7 dataset to conduct our experiments. SpaceNet-7 is a dataset that contains planet satellite imagery mosaics, including 24 images (one per month) covering ≈100 unique geographies of developing area. The dataset comprise over 40,000 square kilometers of imagery and exhaustive polygon labels of building footprints in the imagery, totaling over 10 million individual annotations. We pick each geography's 1st/12th/24th month image, and crop them into 64×64 image sets. We use the knowledge of 1st/12th month data (for original UNet we use 1st month data as input) to predict the 24th month new construction positions. To optimize the neural network, we adopt the RMSprop optimizer with learning rate as 1^{-5}, weight decay as 1^{-8} and momentum as 0.9. We split data as 0.8/0.2 for training and validation. We train the network for 100 epochs.

V. Results and Discussion

Figure 2 shows our results for both the basic and the augmented model. The augmented model outperforms the basic model by 3.8%. While we only report the result for one specific run, we want to anecdotally mention that the result appears to be roughly consistent across multiple runs[1].

We draw two major conclusions from the results in Figure 2:

- First, adding additional timesteps in the form of satellite images appears to improve the performance of the model.
- Second, the model actually also learns from the satellite images and not only from the distorted ground truth we provide as an input (we draw this conclusion since the only additional information the model receives is an additional satellite image).

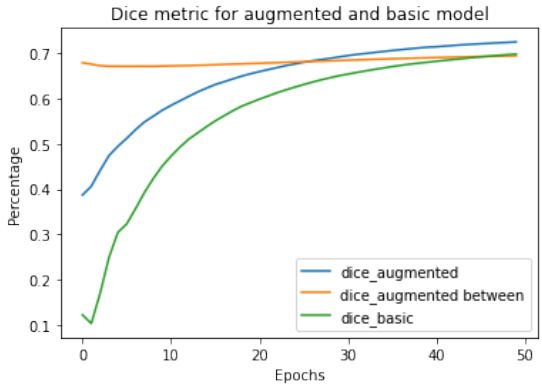

Fig. 2. Dice metric for the two four channel feature map input model ('dice_augmented' with the feature map at two different timescales and 'dice_augmented between' with the duplicated input feature map from 'dice_basic' for validation) and the one feature map input model ('dice_basic'). The x-axis shows the epochs, where epoch 0 denotes the accuracy after the first epoch of training.

In order to further validate our conclusions above, and to exclude the possibility that our augmented model just performs

[1]Future work should report multiple runs with means and standard deviation, however due to the limited scope of this report it is not possible here.

better because of the additional convolutional layers, we also ran the augmented model with an input of two times the input of the basic model. The results are also visible in Figure 2. Interestingly, the model obtains a very high performance very fast (after the first epoch), but does not seem to improve after this. We observed the same behavior running the model multiple times. However, the loss still decreases each epoch similarly to the other models. All in all, it seems like the augmented model can indeed, as concluded above, leverage the additional satellite image.

Problematically, as can be seen in Figure 3, the loss for both models approaches zero very fast. Hence, gradients are very small in later epochs. However, as can be observer, the Sørensen-Dice metric still leaves room for improvement. Currently, our loss function only takes the binary-cross entropy into account. However, as done previously by others (e.g. [4]), incorporating the Sørensen-Dice metric into the loss function could potentially yield further improvements and faster convergence. Additionally, the authors believe that including the Sørensen-Dice metric loss might help to learn without our currently provided ground truth mask. Since, currently, when training without the distorted ground truth mask, the model learns to assign every pixel with the class 'not a building'.

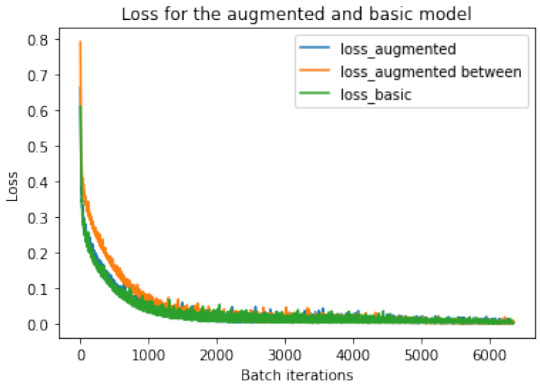

Fig. 3. Loss for the two presented models. 'loss_augmented between' shows the loss for the augmented model, with the two input feature maps being duplicates of the basic models input feature map.

Various outputs of the augmented model at different epochs can be seen in the Appendix (Figure 5).

VI. Conclusion

In this report, we firstly summarize the existing methods for urban planning. Then we propose a novel strategy utilizing the UNet for predicting the urban planning based on the previous urbanization knowledge. However, our method has its limitation. For example, the dataset we adopted doesn't contain the optimized planning since the development in different areas are conducted by different construction companies. Hence, the bias of knowledge is introduced. Also, due to the lack of road map dataset, we cannot predict the development of roads. We hope with more generalizable data, we can further improve our method and make it practical for real-life applications. Since the dataset provides additional granularity in the timesteps than used in this report (only three timesteps) and encouraged by the results in section III we believe that including this additional granularity might lead to even better results. This could be realized with an expanded architecture as presented here, or potentially RNNs, which treat the input images as timeseries data. Another potential avenue could be an adaption of a TGAN [5] - trained on single satellite input images to predict their evolution with time. Furthermore, future work should evaluate how to remove the current dependence on the distorted ground truth. As mentioned previously, we believe this challenge could potentially be (partially) solved by adding the dice loss to the model.

References

[1] Lee R Dice. "Measures of the amount of ecologic association between species". In: *Ecology* 26.3 (1945), pp. 297–302.

[2] Th A Sorensen. "A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons". In: *Biol. Skar.* 5 (1948), pp. 1–34.

[3] Robert Davreu. "Cities of mystery: The lost empire of the Indus Valley". In: *The world's last mysteries* (1978), pp. 121–129.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[5] Masaki Saito, Eiichi Matsumoto, and Shunta Saito. "Temporal generative adversarial nets with singular value clipping". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2830–2839.

[6] Christa Brelsford et al. "Toward cities without slums: Topology and the spatial evolution of neighborhoods". In: *Science advances* 4.8 (2018), eaar4644.

[7] Christa Brelsford, Taylor Martin, and Luıs MA Bettencourt. "Optimal reblocking as a practical tool for neighborhood development". In: *Environment and Planning B: Urban Analytics and City Science* 46.2 (2019), pp. 303–321.

[8] *Datasets*. en. https://spacenet.ai/datasets/. Accessed: 2022-5-3.

APPENDIX

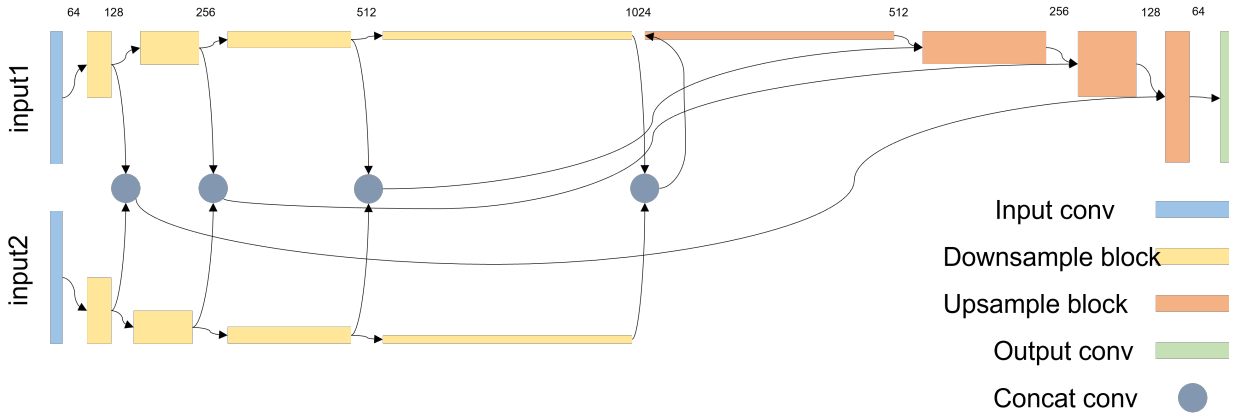

Fig. 4. Our adapted UNet model. Instead of a single input, we provide two four channel input feature maps and downsample them in parallel. Since the upsampling part of the original UNet requires the output at the various downsampling stages, we concatenate and convolve the output of the convolutions at each stage such that the original channel amount is yielded. This convolution is currently performed by a 2×2 filter. Hence, the original upsampling part of UNet can be used. In summary, we added the additional parallel downsample block and the 'Concat conv' blocks.

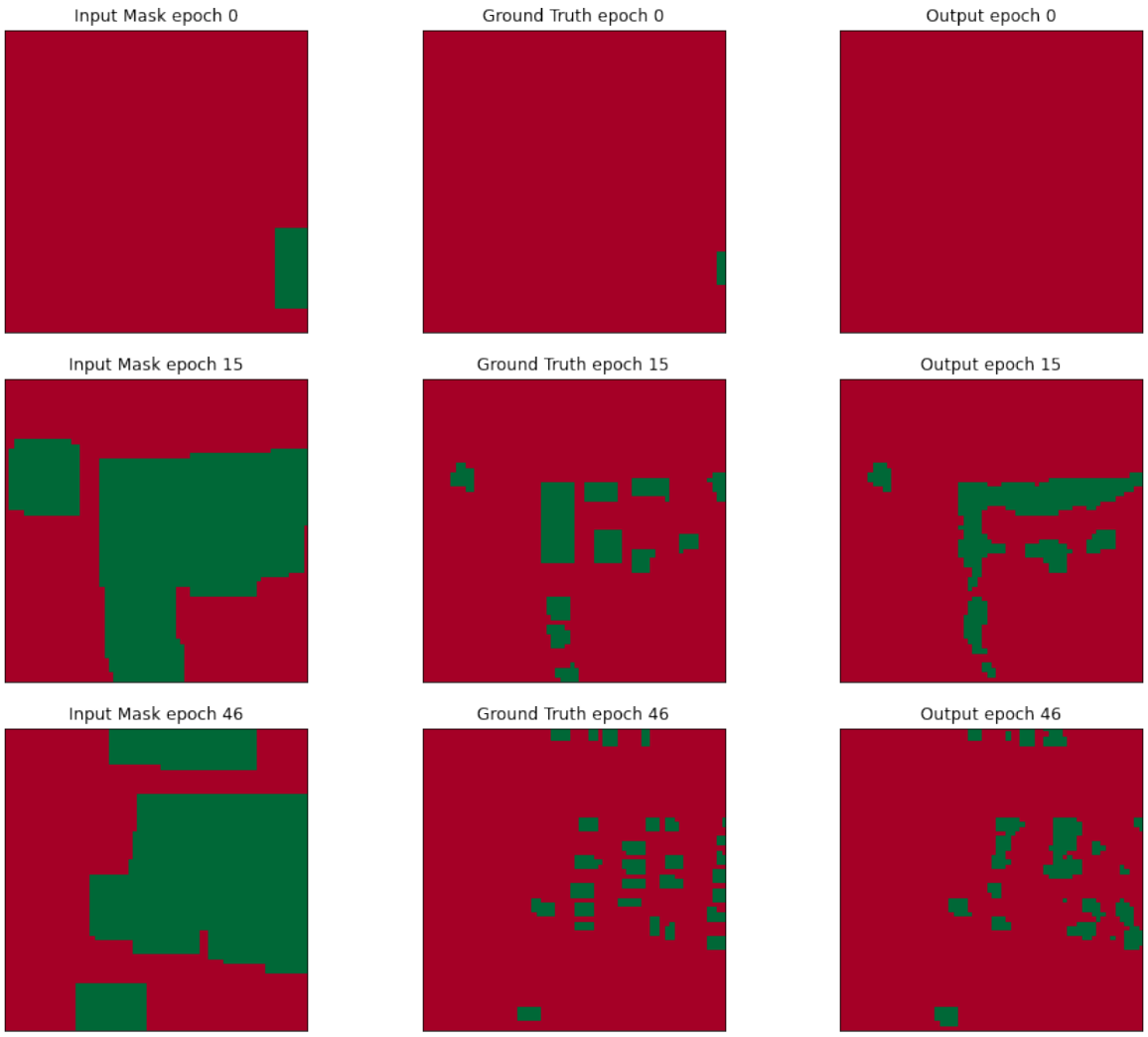

Fig. 5. Outputs, ground truth and input masks for various epochs of the augmented model.