

Traffic4cast at NeurIPS 2021 – Temporal and Spatial Few-Shot Transfer Learning in Gridded Geo-Spatial Processes

<http://traffic4cast.ai> – <https://github.com/iarai/NeurIPS2021-traffic4cast>

Christian Eichenberger*	CHRISTIAN.EICHENBERGER@IARAI.AC.AT
Moritz Neun*	MORITZ.NEUN@IARAI.AC.AT
Henry Martin*†	HENRY.MARTIN@IARAI.AC.AT
Pedro Herruzo*	PEDRO.HERRUZO@IARAI.AC.AT
Markus Spanring*	MARKUS.SPANRING@IARAI.AC.AT
Yichao Lu‡	YICHAO@LAYER6.AI
Sungbin Choi	SUNGBIN.CHOI.1@GMAIL.COM
Vsevolod Konyakhin§, Nina Lukashina¶,	
Aleksei Shpilman §	SEVAKONYAKHIN@GMAIL.COM
Nina Wiedemann†, Martin Raubal†	NWIEDEMANN@ETHZ.CH
Bo Wang**, Hai L. Vu**, Reza Mohajerpoor††, Chen Cai††,	
Inhi Kim‡‡	BO.WANG1@MONASH.EDU
Luca Hermes§§, Andrew Melnik§§, Riza Velioglu§§,	
Markus Vieth§§, Malte Schilling§§	LHERMES@TECHFAK.UNI-BIELEFELD.DE
Alabi Bojesomo ¶¶, Hasan Al Marzouqi¶¶,	
Panos Liatsis¶¶	ALABI.BOJESOMO@KU.AC.AE
Jay Santokhi***, Dylan Hillier***, Yiming Yang***, Joned Sarwar***,	
Anna Jordan***, Emil Hewage***	JAY@ALCHERATECHNOLOGIES.COM
David Jonietz †††	DAVID.JONIETZ@HERE.COM
Fei Tang †††	FEI.TANG@HERE.COM
Aleksandra Gruca†††	ALEKSANDRA.GRUCA@POLSL.PL
Michael Kopp*	MICHAEL.KOPP@IARAI.AC.AT
David Kreil*	DAVID.KREIL@IARAI.AC.AT
Sepp Hochreiter§§§*	SEPP.HOCHREITER@IARAI.AC.AT

Editors: Douwe Kiela, Marco Ciccone, Barbara Caputo

* Institute of Advanced Research in Artificial Intelligence (IARAI), Vienna, Austria

† Institute of Cartography and Geoinformation, ETH Zurich, Switzerland

‡ Layer 6 AI, Toronto, Canada

§ ITMO University, Saint Petersburg, Russia

¶ JetBrains Research, Saint Petersburg, Russia

|| HSE University, Saint Petersburg, Russia

** Institute of Transport Studies, Monash University, Clayton Victoria, Australia

†† CSIRO's Data61, Eveleigh, Australia

‡‡ Institute Civil and Environmental Engineering Department, Kongju National University, South Korea

§§ Machine Learning & Neuroinformatics Group, Bielefeld University, Germany

¶¶ Electrical Engineering and Computer Science Department, Khalifa University, Abu Dhabi, UAE

*** Alchera Data Technologies Ltd, Cambridge, UK

††† HERE Technologies, Zurich, Switzerland

††† Silesian University of Technology, Gliwice, Poland

§§§ Machine Learning Institute, Johannes Kepler University Linz, Austria

Abstract

The IARAI *Traffic4cast* competitions at NeurIPS 2019 and 2020 showed that neural networks can successfully predict future traffic conditions 1 hour into the future on simply aggregated GPS probe data in time and space bins. We thus reinterpreted the challenge of forecasting traffic conditions as a movie completion task. U-Nets proved to be the winning architecture, demonstrating an ability to extract relevant features in this complex real-world geo-spatial process. Building on the previous competitions, *Traffic4cast* 2021 now focuses on the question of model robustness and generalizability across time and space. Moving from one city to an entirely different city, or moving from pre-COVID times to times after COVID hit the world thus introduces a clear domain shift. We thus, for the first time, release data featuring such domain shifts. The competition now covers ten cities over 2 years, providing data compiled from over 10^{12} GPS probe data. Winning solutions captured traffic dynamics sufficiently well to even cope with these complex domain shifts. Surprisingly, this seemed to require only the previous 1h traffic dynamic history and static road graph as input.

1. Introduction

The global trends of urbanization and increased personal mobility force us to rethink the way we use urban space. The *Traffic4cast* competitions (Kreil et al., 2020; Kopp et al., 2021) tackle this problem in a data driven way, encouraging the application of the latest methods in machine learning to modeling complex spatial systems over time.

This year, we provide a unique data set derived from industrial scale trajectories of over 10^{12} raw GPS position fixes, with latitude, longitude, time stamp, as well as vehicle speed and driving direction recorded at that time. The data are made available by HERE Technologies and originate from a large fleet of vehicles. For the new temporal and spatial transfer learning challenges introduced in *Traffic4cast* 2021, we provide data for 10 culturally and socially diverse metropolitan areas around the world, covering a time span of 2 years.

The competition task was to predict from 1 hour traffic the next 5, 10, 15, 30, 45 and 60 min into the future. The training data was provided in the same format as last year (Kopp et al., 2021). An overview of the dynamic training data provided and the 1h test time slots to predict is shown in Figure 1. Along with each 1h test time slot, we provided the time of day and the day of the week but not the exact date. For each city, we provided a static graph derived from a road map in the same spatial resolution as the dynamic data, which could be used as mask or as a graph (Eichenberger et al., 2022; Eichenberger and Neun, 2021).

For all cities considered, COVID-19 has lead to a visible shift in daily mobility traffic patterns, as can be seen by comparing traffic volumes from the pre-Covid (April/May 2019) and in-Covid data (April/May 2020), see also Eichenberger et al. (2022) and Eichenberger and Neun (2021). Our **core challenge** requires participants to transfer learn across this domain shift (Ben-David et al., 2010; Wouter, 2018; Kouw and Loog, 2019; Webb et al., 2018; Gama et al., 2014; Widmer and Kubat, 1996) in traffic caused by the COVID-19 pandemic. This challenge is thus a **few-shot learning task** (Fei-Fei et al., 2006; Lu et al., 2020; Guo et al., 2019), which requires to **transfer learn traffic dynamics across a temporal domain shift**.

Our **extended challenge** encourages participants to use all the data provided so far (data from 8 different cities for the pre-COVID era and 4 different cities during the in-COVID

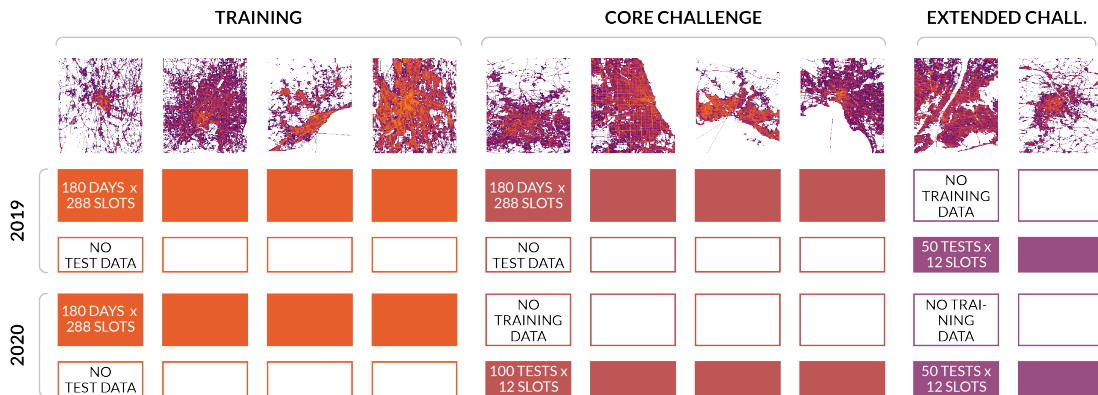


Figure 1: Data overview. There is data from 10 cities from 2019 and 2020. 4 cities are used for training (180 full days 2019 and 2020; 4 cities are used for the core challenge (180 full days 2019 pre-Covid and 100 test slots 2020 in-Covid); 2 cities are used for the extended challenge (50 test slots 2019 pre-Covid and 50 test slots 2020 in-Covid).

era) on two hitherto unseen cities for which no further training data is provided. For each city, 100 one-hour test time slots are randomly chosen, 50 from the pre-COVID era and 50 from the in-COVID era. The underlying machine learning challenge is thus a **few-shot transfer of traffic dynamics across a spatial and temporal domain shift**. It is noted that solutions to the extended challenge can be used as a solution for the core challenge as well, although participants will then make no use of the additional training data which might contain local, spatially relevant information. Apart from the static data and the test time slots, no further data from the target cities are exposed.

Informed by the previous *Traffic4cast* competitions (Kreil et al., 2020; Kopp et al., 2021), we chose two non-trivial baselines. For the core competition, we used a vanilla U-Net (Ronneberger et al., 2015) and trained models for each city separately for 4 epochs on the city’s 2019 training data only. For the extended competition, a Graph ResNet was used following Martin et al. (2020). More details about the baselines can be found in Eichenberger et al. (2022) and Eichenberger and Neun (2021).

The competition brings together a range of highly active fields in machine learning – few-shot learning, transfer learning, meta-learning more generally, as well as video frame prediction or graph based modelling. Compared to last year’s competition an order of magnitude more data was provided, covering ten cities across 2019 and 2020. This wealth of data was the basis for being able to tackle how far data and machine learning driven approaches alone can be used to decipher the implicit, largely unknown rules governing the phenomena of traffic by applying them across complex domain shifts.

The data encoding as Traffic Map Movies (Kreil et al., 2020), featuring multiple channels, provides a natural way to fuse information from multiple sources and show-cases the power of Machine Learning to excel at tasks that previously had to depend on domain knowledge, special data structures working on graphs (Snyder and Do, 2019), and manual feature engineering. Leading the way, this approach also proved itself in other domains such as

rainfall prediction (Agrawal et al., 2019; Gruca et al., 2021; Herruzo et al., 2021), featured at CIKM and IEEE Big Data 2021 (<https://weather4cast.ai/>).

2. Standout Solutions

In the following, we give a short summary of outstanding solutions to our Traffic4cast 2021 competition. For each submission, we give a high-level diagram which we call inference diagram and which, in contrast to architecture diagrams typically used in the ML literature, summarizes the approach from an information-flow perspective, highlighting the trained models, the data used to train these models and to the ensembling of these trained models at inference time (if any). This sheds light on how the large and diverse amount of input data was used for the different tasks. The notation is the following: rectangles refer to data (test input, and test output); rounded rectangles represent functions in the inferences, arrows represent flow of information. We use square brackets to denote parameterized data or functions, bound to the parameters in the test input. For each model, after the colon, we also show the data the model was trained on, using the notation in Table 1, see ^d in caption.

2.1. oahciy: U-Net + Multi-Task Learning

Lu (2021) presents an amazingly simple multi-task learning framework by randomly sampling data from all available cities (4 training and 4 core) for all models in the ensemble. In the core competition 9 models were trained for 5 epochs while in the extended competition 7 models were trained for 50'000 steps only (8% of an epoch). The models are all U-Net with varying architecture and seeds. For more details see Table 1 and Appendix A.1.

Lu (2021) argues that this multi-task learning can be regarded as an implicit data augmentation and regularization technique when trained on one city only and forcing to learn city-agnostic representations thereby improving data efficiency. The implicit domain-adaptation through the addition of 2019 and 2020 data for at least one city is reported as crucial encouraging the model to learn to adapt to temporal domain shifts during training.

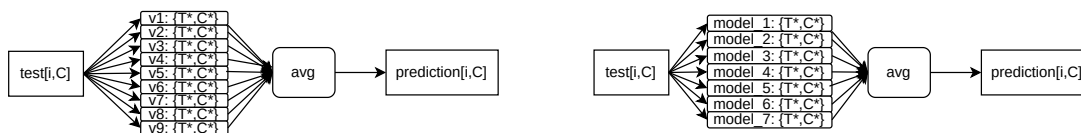


Figure 2: Inference oahciy (Lu, 2021) (left: core competition, right: extended competition).

2.2. sungbin: U-Net Ensemble

The approach of Choi (2021) is very similar to (Lu, 2021), also averaging ensembles of different U-Net architectures (4 city-independent models). In contrast to Lu (2021), Choi (2021) trained on target city training data, too (3 models in core). For more details see Table 1 and Appendix A.2.

2.3. sevakon: U-Net with Temporal Domain Adaptation

Konyakhin et al. (2021) also base their approach on the success of U-Nets in previous competitions. However, in contrast to (Lu, 2021) and (Choi, 2021), they train their models

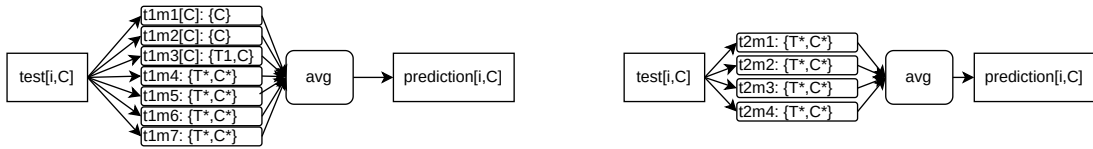


Figure 3: Inference sungbin (Choi, 2021) (left: core competition, right: extended competition).

on each target city in the core competition only (they did not participate in the extended competition). They use three different architectures (vanilla U-Net, DenseNet, and EfficientNet (Tan and Le, 2019) pre-trained on Imagenet (Deng et al., 2009)), a static mask derived from dynamic data and a per-pixel and per-channel temporal domain-adaptation (TDA) factor. Their final prediction is derived from these 3 models; each model is used with and without TDA, resulting in 6 predictions to which the static mask is applied and which are then averaged.



Figure 4: Inference sevakon (Konyakhin et al., 2021) (core competition). The city mask is derived from the training and test data.

2.4. nina: U-Net++ on Patches

Wiedemann and Raubal (2021) also use a U-Net variant, but in a patch-based manner, as it was shown to be beneficial in other segmentation tasks (Zhang et al., 2006; Ghimire et al., 2020) (also (Misra et al., 2020) in classification). No static road information was used. This method allowed them to use a parameter-heavier UNet++ with more skip connections (Zhou et al., 2019, 2018), which they suggest was helpful in light of the sparsity of the data. Patches are sampled from all available labelled data. After different experiments, choosing the available patches to be 100×100 crops with stride 10 was found to perform best. The patch-wise prediction exhibits an ensemble-like behavior.

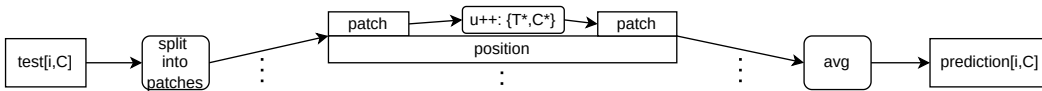


Figure 5: Inference nina (Wiedemann and Raubal, 2021) (core competition and extended competition).

2.5. ai4ex: SWIN-Transformer

Bojesomo et al. (2021) uses a Swin-UNet structure where all convolution blocks are replaced by shifted window self attention; downsampling in the encoder is achieved by trainable patch merging layers and upsampling by patch expanding layers in the decoder branch; skip connections are implemented by a combination of addition and concatenation.

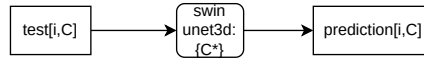


Figure 6: Inference ai4ex (Bojesomo et al., 2021) (core and extended competition).

2.6. dninja: Graph-Based U-Net

Hermes et al. (2022) present a graph based approach aiming at better generalization and transfer by leveraging knowledge of the underlying road network whilst ignoring areas without any traffic information. In order for their solution to make full use of the 2d topological information contained in the competition data they use 4 subgraphs corresponding to the provided 4 heading channels of said data.

2.7. resuly: 3DResNet and Sparse-UNet

Wang et al. (2021a) use a 3DResnet (Wang et al., 2021b) with 3D convolutions in 4 residual blocks and an output block. For the core competition, this output block consists of sequential CNN layers to restrain the temporal relationship. For the extended competition, the output block consists of a sparse U-Net (Graham, 2014; Choy et al., 2019) with data in Coordinate Format (COO) for the extended competition.

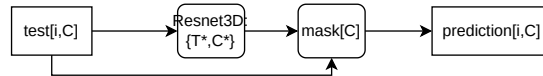


Figure 7: Inference resuly (Wang et al., 2021a) (left: core competition, right: extended competition). The mask is derived from the test data.

2.8. jaysantokhi: Dual-Encoding U-Net

Santokhi et al. (2021) use a dual encoding U-Net architecture aiming at a lightweight approach for real world deployments containing significantly fewer parameters (see also Table 1) and shorter training times. The architecture consists of two encoders one of which has skip connections to the decoder; encoder and decoder consist of Convolutional LSTM layers. The skip connections are not vanilla, but designed to carry the hidden and cell states of the encoder LSTM to the decoder LSTM, which is crucial for the approach. In both competitions, 4 models are pre-trained on the training cities and fine-tuned on the core competition cities. In the core competition, the city-specific fine-tuned model is used, whereas in the extended competition an architecture with fewer parameters is used and predictions are built by averaging over the outputs of all 4 models.

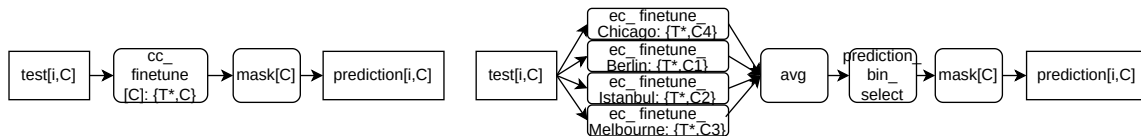


Figure 8: Inference jaysantokhi (Santokhi et al., 2021) (left: core competition, right: extended competition). The city mask is derived from test data.

3. Synopsis and Discussion

Looking at the different solutions above we see a large variety of mostly U-Net based approaches. Based on the experiences of the previous year (Kopp et al., 2021), this is not a complete surprise. However, it is interesting to see that for instance the simple averaging in ensembles in the winning solutions was also able to handle the domain shifts, even slightly outperforming some more tailored domain adaptation techniques. Table 1 highlights the key aspects and differences of the chosen architectures and informs the discussion below.

Team, rank (core/ext), approach	road graph, time-of-day, day-of-week ^a	models trained p. city ^b	#models trained ^c	Training datasets ^d	\sum #params core / ext ^e	mask ^f
oahciy (1/2) U-Net + multi-task learning (Lu, 2021)	road graph (concat)	no	9 / 7	$(9/7) \times \{T^*, C^*\}$	1710.2M / 17.1M	–
sungbin (2/1) U-Net Ensemble (Choi, 2021)	road graph (concat)	in two U-Nets	16/4	$2 \times \{C[1-4]\}; \{T1, C[1-4]\}; 4 \times \{T^*, C^*\} / 4 \times \{T^*, C^*\}$	123.6M / 33.9M	–
sevakon (3/-) U-Net with Temporal Domain Adaptation (Konyakhin et al., 2021)	no	yes	11/-	$3 \times \{C1\}; 2 \times \{C2\}; 3 \times \{C3\}; 3 \times \{C4\}$	342.0M / –	city (train/test data)
nina (6/3) U-Net++ on patches (Wiedemann and Raubal, 2021)	no	no	1=1	$\{T^*, C^*\}$	36.7M / 36.7M	–
ai4ex (4/6) SWIN-Transformer (Bojesomo et al., 2021)	no	no	1=1	$\{C^*\}$	141.9M / 141.9M	–
dninja (7/4) Graph-based U-Net (Hermes et al., 2022)	road graph, time-of-day, day-of-week	no	1=1	$\{T^*, C^*\}$	5.8M / 5.8M	by GNN
resuly (5/-) 3DResNet, Sparse-UNet (Wang et al., 2021a)	road graph	no	1/1	$\{T^*, C^*\}$	17.3M / 43k	test (test data)
jaysantokhi (8/5) Dual-Encoding U-Net (Santokhi et al., 2021)	no	after pre-training	4/4	$\{T^*, C[1-4]\}$	1.0M / 0.3M	city (test data)

Table 1: Synopsis. ^a what supplemental information is used; ^b whether some of the trained models used in the inference are specifically trained on the target city; ^c 9/7 means 9 models in the core and 7 different models in the extended competition, whereas 1=1 means the same trained model was used in both competition; ^d T1=Moscow, T2=Barcelona, T3=Antwerp, T4=Bangkok, C1=Berlin, C2=Istanbul, C3=Melbourne, C4=Chicago, E1=Vienna, E2=New York, T*=all training cities, C*=all core cities, E*=all extended cities. E.g. $(9/7) \times \{T^*, C^*\}$ means 9 models trained on all training and core competition cities for the core competition and 7 from the same cities for the extended competition, and $\{T^*, C[1-4]\}$ expands to a model for each city trained on all training cities plus one of the core cities; ^e Sum of trainable weights of all the model checkpoints used in the inference as extracted from the participants’ checkpoint using code in Eichenberger and Neun (2021); ^f Kind of mask used for post-processing.

3.1. Why were the same strategies successful in both competitions?

Lu (2021) won the core competition with an ensemble of models all trained with data from all training and core competition cities. This indicates that his solution already is able to capture spatial transfer, although at the price of a high amount of parameters (see Table 1). Hence, the combination of a diverse enough set of training cities with a large amount of data together with the static road information seems to be enough to solve both transfer learning challenges of our competition. Although the second placed approach of (Choi, 2021) is very similar, apart from minor architectural choices, the main difference is that Lu (2021) only has city-independent models. In contrast, the models by Konyakhin et al. (2021) trained only per city are competitive with regards to the temporal transfer. This again gives an indication that the explicit temporal domain-adaptation was necessary in this approach. In contrast, the first two ensembles (Lu et al., 2020; Choi, 2021) successfully did an implicit temporal adaptation through the city-independent models trained on data from multiple cities from before and after the temporal shift. Naively, we would have expected to see more domain-adaptation approaches in the core competition like the temporal domain-adaptations by (Konyakhin et al., 2021) or data augmentation techniques to use the test slot inputs for few-shot learning.

It is also remarkable to see that the patch-based approach of Wiedemann and Raubal (2021) is competitive especially for the spatial transfer in the extended challenge. The patches introduced an additional implicit level of ensemble learning within a city. This seems to have had a similar positive impact. The work of Bojesomo et al. (2021), Hermes et al. (2022) and Santokhi et al. (2021) show that transformers on patches, graph-based approaches and light-weight UNets are also able to handle the transfer tasks in both competitions with, in some cases, significantly smaller models (see Table 1).

In addition, we see clear traces of temporal averaging in the qualitative analysis of the outlier special prize (see Section 3.4 and Appendix C). Hence, it seems that non-distributional predictions evaluated by MSE seem to encourage trading off spatial, temporal, and channel-wise features jointly. A similar temporal and spatial “blurring” effect has been reported in weather forecasting (Ravuri et al., 2021b; Sønderby et al., 2020; Witt et al., 2021).

Finally, we can see that most approaches did not use any additional dynamic features such as time-of-day or day-of-week. Hence, 1h dynamic input data together with static data seems to capture the city-specific dynamics already, which models can and did then exploit in both our domain shift transfer learning challenges. Of course, one cannot exclude that completely different approaches than those considered and submitted by our competition participants could benefit from such additional features or that they could be beneficial in a different task setting (see Section 3.5).

3.2. Where do the submission performances differ?

We analyse the MSE loss for each solution above per directed cell (viewing all 4 headings of a pixel as virtual detector for volume and speed separately) – we bin these directional pixels by their standard deviation in the ground truth speed data (details and full analysis can be found in Appendix B). Thereby, we see where predictions are hard and where the competition was decided. We see that all models struggle in the same std ranges. Of course, MSE does not optimize each directed location independently, so the interpretation here has to be taken

cum grano salis. Referring to Figure 9, most of speed MSE losses for all solutions under

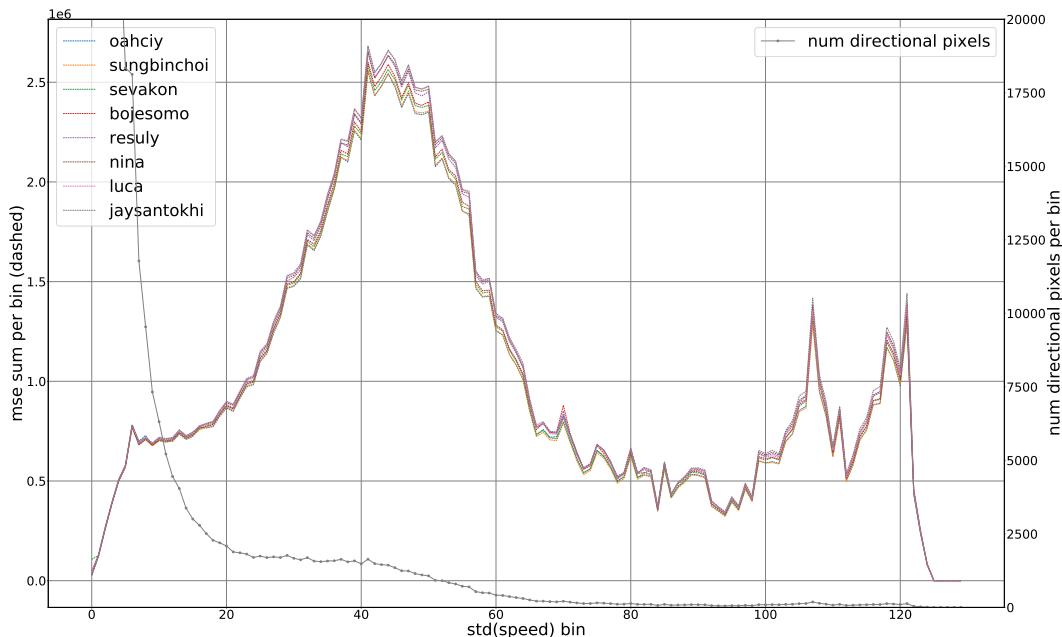


Figure 9: Relating MSE to std for speeds in BERLIN core: distribution of std among oriented pixels and summed MSE. The shaded gray areas highlights the two critical speed std ranges.

consideration come from the ranges 35–55 and 105–125 in speed std. If no data is collected, speed is set to 0, hence, the higher range implies an oscillation between very high speeds in the upper part of the value range $[0, 255]$ and no-data and all solutions seem to have struggled in providing accurate predictions. We localized these two ranges on the city map, see Appendix B, and observe: 35–55 speed std range tends to cover the main arteria and some long-range country roads with commuting traffic starting early in the morning; the 105–125 speed std range tends to cover areas with usually high speeds. This would hint at all solutions struggling where it is hard to find a good strategy of predicting speed in areas of very high variance and in the large area of medium standard deviation where the amount of such locations makes the total penalty so high.

3.3. What is the contribution of static road information?

Regarding static road information, Lu (2021) and Choi (2021) did not explicitly compare training with and without static road information as input. The patch-based approach of Wiedemann and Raubal (2021) does not use the static information; this might explain why this approach ranks lower in the core competition. We suppose that for many test inputs, the static road information is already in the dynamic data; however, in particular in areas of data sparsity, the static road information may be essential. So it remains unclear to what degree static information is required and, in particular, whether it helps in the critical ranges above.

3.4. What did models learn in outlier situations?

In order to tackle this questions, we asked participants to re-run their models on a new test set provided for our Outlier Special Prize. The test slots were from two cities of the core competition and the slots were quantitatively evaluated with a masked MSE. For the Special Prize, that mask only contained one pixel and two channels (hence $200 \cdot 6 \cdot 2$ integer values were evaluated against ground truth). These pixels were selected to contain a traffic jam situation reflected by a drastic drop in speed with a simultaneous increase in volume. In the examples analysed in Appendix C, the winning models of our core and extended competition predict a smoothed version of a jam resolution, underestimating speed and overestimating density. As MSE encourages model predictions to tend towards the mean in the data, they are never predicting a rare scenario such as a jam resolving more or less quickly than in expectation. In the selected outliers MSE for volume and speed is at the same level. More details on the heuristics and the design of our Outlier Special Prize can be found in Appendix C.

3.5. What did we learn about the metric?

Both challenges used the pixel-wise mean squared error (MSE) as a loss metric as it is also used in movie prediction tasks (Srivastava et al., 2015; Lee et al., 2018; Kwon and Park, 2019; Walker et al., 2016; Xue et al., 2016; Han et al., 2019; Oprea et al., 2020). From the analysis in previous subsections pertaining to this year’s competition challenges as well as from similar observations from our previous competitions (Kreil et al., 2020; Kopp et al., 2021), we are able to identify important, desirable properties a loss metric should have without currently being able to construct one. These properties are as follows.

Desired Property 1: The metric should not be distorted by missing values. Zero speed values in the case of no-data lead to a high variance in the speeds to predict in any temporally gridded bin. In contrast, zero volume for missing values does not affect the variance. A possible solution, suggested by Wiedemann and Raubal (2021), is to evaluate speed only in case of non-zero volume. With that models might be less defensive predicting speed.

Desired Property 2: The metric should be flexible in handling spatial and temporal shifts. Predicting just slightly at the wrong place leads to a double punishment, encouraging solutions that can be visually observed to show an ‘averaging phenomenon’ even in high scoring solutions as outlined in Qi and Kwok (2020). Shifts in traffic phenomena need to be handled depending on the application needs.

Desired Property 3: The metric should cope with different scales. Volume values can be heavily biased due to the fact that GPS probes are only collected by a fraction of the full traffic. Speed values are in most cases less biased (apart from idiosyncrasies such as (un)loading) but often cluster (e.g. around signaled speed limits under normal traffic conditions). In the competition no normalization was used on purpose, giving more weight to speed predictions. Furthermore, outlier situations are rare, but critical, and therefore tend to not be given enough weight compared to their utility by MSE. Hence, the metric should be able to incorporate multiple aggregation levels.

Desired Property 4: The metric should be distribution-aware. Single-value predictions tend to produce averaging scenarios as we see in the analysis of outliers (see Appendix C). Similar to Espeholt et al. (2021) and Ravuri et al. (2021a) predicting a possible distribution

instead of a single scalar could help estimating the uncertainty and evaluating at the distributional level. Such an approach would probably make most sense in combination with *Desired Property 2*.

3.6. Has transfer learning been achieved? In what sense?

If we think of the winning U-Net architecture (Lu, 2021), then models are able to combine global and local information to create a good average forecast. However, the pixel-wise MSE metric leads to models that do not yet cover sudden changes over time that can occur in real-world traffic dynamics (see Sections 3.4 and 3.5).

4. Summary and Outlook

It is encouraging to see that so many different machine learning approaches lead to competitive results in our map movie completion task. Winning solutions captured traffic dynamics sufficiently well to even cope with complex domain shifts. Surprisingly, this seemed to require only the previous 1h traffic dynamic history and static road graph as input. In addition, our competition results point to interesting future research directions and questions.

The pixel-wise MSE loss metric encouraged solutions that performed poorly on real-world relevant outlier situations with their prediction often being averaged scenarios. Improving on this metric may necessitate a range of relevant real-world tasks, each with their own external task-specific metric and reference data, such as predicting estimated times of arrival, ETA (Derrow-Pinion et al., 2021). In another domain, the climate extremes indices provide a standard catalogue of metrics (Alexander et al., 2021). The need to consider complementary domain specific tasks was recently also demonstrated in precipitation forecasting (Ravuri et al., 2021a; Espeholt et al., 2021). For traffic, unfortunately, there seems to be no such standard catalogue of application tasks, metrics, and data.

The *Traffic4cast* 2021 competition has proven again that representing traffic in a temporal and spatial grid is a powerful framework that allows for complex questions in traffic research to be addressed. The prediction of effects of road closures, changed speed limits, or an addition of lanes could thus be equally formulated in such a representation as simple completion tasks. Fundamental questions of classical traffic research can thus be approached in a purely data-centric, non-expert dependent way with the help of machine learning. Finally, it is of key interest to explore whether this framework can be adapted or extended to incorporate new data sources such as satellite data, loop counter data, or crowd-sourced traffic data. The necessary sensor fusion and cross-platform translation tasks are interesting applications of machine learning.

Acknowledgments

We would like to thank HERE technologies for making our competition data available.

References

- Shreya Agrawal, Luke Barrington, Carla Bromberg, John Burge, Cenk Gazen, and Jason Hickey. Machine learning for precipitation nowcasting from radar images. arXiv preprint arXiv:1912.12132, 2019. URL <https://arxiv.org/abs/1912.12132>.
- Lisa Alexander, Markus Donat, Margot Bador, Ian Macadam, Nicholas Herold, and Jonathan McComb. Climdex indices – the climdex indices can help us understand patterns in temperature and precipitation extremes: how they change from year to year or from place to place., 2021. URL <https://www.climdex.org/learn/indices/>. Accessed 2022-02-11.
- S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. Vaughan. A theory of learning from different domains. Machine Learning, 79:151–175, 2010.
- Alabi Bojesomo, Hasan Al-Marzouqi, and Panos Liatsis. Swinunet3d – a hierarchical architecture for deep traffic prediction using shifted window transformers. arXiv preprint arXiv:2201.06390v1, 2021. URL <https://arxiv.org/abs/2201.06390v1>.
- Sungbin Choi. Applying unet for the traffic map prediction across different time and space, 2021. URL <https://github.com/SungbinChoi/traffic4cast2021/blob/main/Paper.pdf>.
- Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3075–3084, 2019.
- Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.
- Austin Derrow-Pinion, Jennifer She, David Wong, Oliver Lange, Todd Hester, Luis Perez, Marc Nunkesser, Seongjae Lee, Xueying Guo, Brett Wiltshire, and et al. Eta prediction with graph neural networks in google maps. Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Oct 2021. doi: 10.1145/3459637.3481916. URL <http://dx.doi.org/10.1145/3459637.3481916>.
- Christian Eichenberger and Moritz Neun. Traffic4cast 2021 – temporal and spatial few-shot transfer learning in traffic map movie forecasting. <https://github.com/iarai/NeurIPS2021-traffic4cast>, July 2021. Accessed 2022-01-28.
- Christian Eichenberger, Moritz Neun, Henry Martin, Pedro Herruzo, Markus Spanring, Yichao Lu, Sungbin Choi, Vsevolod Konyakhin, Nina Lukashina, Aleksei Shpilman, Nina Wiedemann, Martin Raubal, Bo Wang, Hai L. Vu, Reza Mohajerpoor, Chen Cai, Inhi Kim, Luca Hermes, Andrew Melnik, Riza Velioglu, Markus Vieth, Malte Schilling, Alabi Bojesomo, Hasan Al Marzouqi, Panos Liatsis, Jay Santokhi, Dylan Hillier, Yiming Yang, Joned Sarwar, Anna Jordan, Emil Hewage, David Jonietz, Fei Tang, Aleksandra Gruca, Michael Kopp, David Kreil, and Sepp Hochreiter. Traffic4cast at neurips 2021 – temporal and spatial few-shot transfer learning in gridded geo-spatial

- processes. arXiv preprint arXiv:2203.17070, 2022. doi: 10.48550/ARXIV.2203.17070. URL <https://arxiv.org/abs/2203.17070>. Preprint with Supplementary Material.
- Lasse Espeholt, Shreya Agrawal, Casper Sønderby, Manoj Kumar, Jonathan Heek, Carla Bromberg, Cenk Gizen, Jason Hickey, Aaron Bell, and Nal Kalchbrenner. Skillful twelve hour precipitation forecasts using large context neural networks. arXiv preprint arXiv:2111.07470, 2021. URL <https://arxiv.org/abs/2111.07470>.
- L. Fei-Fei, R. Fergus, and P. Perona. One-shot learning of object categories. IEEE transactions on pattern analysis and machine intelligence, 28(4):594–611, 2006.
- J. Gama, I. Zliobaite, A. Bifet, P. Mykola, and A. Bouchachia. A survey on concept drift adaptation. ACM Computing Surveys, 46(4), 2014. ISSN 0360-0300. doi: 10.1145/2523813.
- Kanchan Ghimire, Quan Chen, and Xue Feng. Patch-based 3d unet for head and neck tumor segmentation with an ensemble of conventional and dilated convolutions. In 3D Head and Neck Tumor Segmentation in PET/CT Challenge, pages 78–84. Springer, 2020.
- Benjamin Graham. Spatially-sparse convolutional neural networks. arXiv preprint arXiv:1409.6070, 2014. URL <https://arxiv.org/abs/1409.6070>.
- Aleksandra Gruca, Pedro Herruzo, Pilar Ripodas, Andrzej Kucik, Christian Briese, Michael K. Kopp, Sepp Hochreiter, Pedram Ghamisi, and David P. Kreil. CDCEO’21 - First Workshop on Complex Data Challenges in Earth Observation, page 4878–4879. Association for Computing Machinery, New York, NY, USA, 2021.
- Y. Guo, N. Codella, L. Karlinsky, J.R. Smith, T. Rosing, and R. Feris. A new benchmark for evaluation of cross-domain few-shot learning. arXiv preprint arXiv:1912.07200, 2019. URL <https://arxiv.org/abs/1912.07200>.
- Tengda Han, Weidi Xie, and Andrew Zisserman. Video representation learning by dense predictive coding. In Proceedings of the IEEE International Conference on Computer Vision Workshops, pages 0–0, 2019.
- Luca Hermes, Andrew Melnik, Riza Velioglu, Markus Vieth, and Malte Schilling. A graph-based u-net model for predicting traffic in unseen cities. arXiv preprint arXiv:2202.06725, 2022. URL <https://arxiv.org/abs/2202.06725>.
- Pedro Herruzo, Aleksandra Gruca, Llorenç Lliso, Xavier Calbet, Pilar Rípodas, Sepp Hochreiter, Michael Kopp, and David P. Kreil. High-resolution multi-channel weather forecasting – First insights on transfer learning from the Weather4cast Competitions 2021. In 2021 IEEE International Conference on Big Data (Big Data), pages 5750–5757, December 2021. doi: 10.1109/BigData52589.2021.9672063.
- Vsevolod Konyakhin, Nina Lukashina, and Aleksei Shpilman. Solving traffic4cast competition with u-net and temporal domain adaptation. arXiv preprint arXiv:2111.03421, 2021. URL <https://arxiv.org/abs/2111.03421>.

- Michael Kopp, David Kreil, Moritz Neun, David Jonietz, Henry Martin, Pedro Herruzo, Aleksandra Gruca, Ali Soleymani, Fanyou Wu, Yang Liu, Jingwei Xu, Jianjin Zhang, Jay Santokhi, Alabi Bojesomo, Hasan Al Marzouqi, Panos Liatsis, Pak Hay Kwok, Qi Qi, and Sepp Hochreiter. Traffic4cast at neurips 2020 – yet more on the unreasonable effectiveness of gridded geo-spatial processes. In Hugo Jair Escalante and Katja Hofmann, editors, Proceedings of the NeurIPS 2020 Competition and Demonstration Track, volume 133 of Proceedings of Machine Learning Research, pages 325–343. PMLR, 2021. URL <http://proceedings.mlr.press/v133/kopp21a.html>.
- W. M. Kouw and M. Loog. A review of domain adaptation without target labels. IEEE transactions on pattern analysis and machine intelligence, October 2019. ISSN 0162-8828. doi: 10.1109/tpami.2019.2945942.
- David P Kreil, Michael K Kopp, David Jonietz, Moritz Neun, Aleksandra Gruca, Pedro Herruzo, Henry Martin, Ali Soleymani, and Sepp Hochreiter. The surprising efficiency of framing geo-spatial time series forecasting as a video prediction task – insights from the iarai *Traffic4cast* competition at neurips 2019. In Hugo Jair Escalante and Raia Hadsell, editors, Proceedings of the NeurIPS 2019 Competition and Demonstration Track, volume 123 of Proceedings of Machine Learning Research, pages 232–241. PMLR, 08–14 Dec 2020. URL <http://proceedings.mlr.press/v123/kreil20a.html>.
- Yong-Hoon Kwon and Min-Gyu Park. Predicting future frames using retrospective cycle gan. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1811–1820, 2019.
- Alex X Lee, Richard Zhang, Frederik Ebert, Pieter Abbeel, Chelsea Finn, and Sergey Levine. Stochastic adversarial video prediction. arXiv preprint arXiv:1804.01523, 2018. URL <https://arxiv.org/abs/1804.01523>.
- J. Lu, P. Gong, J. Ye, and C. Zhang. Learning from very few samples: A survey, 2020.
- Yichao Lu. Learning to transfer for traffic forecasting via multi-task learning. arXiv preprint arXiv:2111.15542, 2021. URL <https://arxiv.org/abs/2111.15542>.
- Henry Martin, Dominik Bucher, Ye Hong, René Buffat, Christian Rupprecht, and Martin Raubal. Graph-ResNets for short-term traffic forecasts in almost unknown cities. In NeurIPS 2019 Competition and Demonstration Track, pages 153–163. PMLR, August 2020. URL <http://proceedings.mlr.press/v123/martin20a.html>. ISSN: 2640-3498.
- Debaleena Misra, Carlos F Crispim-Junior, and Laure Tougne. Patch-based CNN evaluation for bark classification. In Workshop on Computer Vision Problems in Plant Phenotyping, Edinburgh, United Kingdom, August 2020. URL <https://hal.archives-ouvertes.fr/hal-02969811>.
- Sergiu Oprea, Pablo Martinez-Gonzalez, Alberto Garcia-Garcia, John Alejandro Castro-Vargas, Sergio Orts-Escolano, Jose Garcia-Rodriguez, and Antonis Argyros. A review on deep learning techniques for video prediction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

- Qi Qi and Pak Hay Kwok. Traffic4cast 2020—graph ensemble net and the importance of feature and loss function design for traffic prediction. arXiv preprint arXiv:2012.02115, 2020. URL <https://arxiv.org/abs/2012.02115>.
- Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, and et al. Skilful precipitation nowcasting using deep generative models of radar. Nature, 597(7878):672–677, Sep 2021a. ISSN 1476-4687. doi: 10.1038/s41586-021-03854-z. URL <http://dx.doi.org/10.1038/s41586-021-03854-z>.
- Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, Rachel Prudden, Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Niall Robinson, Ellen Clancy, Alberto Arribas, and Shakir Mohamed. Skilful precipitation nowcasting using deep generative models of radar. Nature, 597(7878):672–677, September 2021b. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03854-z. URL <https://www.nature.com/articles/s41586-021-03854-z>.
- 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, pages 234–241. Springer, 2015.
- Jay Santokhi, Dylan Hillier, Yiming Yang, Joned Sarwar, Anna Jordan, and Emil Hewage. Dual encoding u-net for spatio-temporal domain shift frame prediction. arXiv preprint arXiv:2110.11140, 2021. URL <https://arxiv.org/abs/2110.11140>.
- Corey Snyder and Minh Do. Streets: A novel camera network dataset for traffic flow. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, Advances in Neural Information Processing Systems, volume 32. Curran Associates, Inc., 2019. URL <https://proceedings.neurips.cc/paper/2019/file/ee389847678a3a9d1ce9e4ca69200d06-Paper.pdf>.
- Nitish Srivastava, Elman Mansimov, and Ruslan Salakhudinov. Unsupervised learning of video representations using lstms. In International conference on machine learning, pages 843–852, 2015.
- Casper Kaae Sønderby, Lasse Espeholt, Jonathan Heek, Mostafa Dehghani, Avital Oliver, Tim Salimans, Shreya Agrawal, Jason Hickey, and Nal Kalchbrenner. MetNet: A Neural Weather Model for Precipitation Forecasting. arXiv:2003.12140 [physics, stat], March 2020. URL <http://arxiv.org/abs/2003.12140>. arXiv: 2003.12140.
- Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning, pages 6105–6114. PMLR, 2019.
- Jacob Walker, Carl Doersch, Abhinav Gupta, and Martial Hebert. An uncertain future: Forecasting from static images using variational autoencoders. In European Conference on Computer Vision, pages 835–851. Springer, 2016.

- Bo Wang, Reza Mohajerpoor, Chen Cai, Inhi Kim, and Hai L. Vu. Traffic4cast – large-scale traffic prediction using 3dresnet and sparse-unet. arXiv preprint arXiv:2111.05990, 2021a. URL <https://arxiv.org/abs/2111.05990>.
- Bo Wang, Hai L Vu, Inhi Kim, and Chen Cai. Short-term traffic flow prediction in bike-sharing networks. Journal of Intelligent Transportation Systems, pages 1–18, 2021b. ISSN 1547-2450.
- G.I. Webb, L.K. Lee, B. Goethals, and F. Petitjean. Analyzing concept drift and shift from sample data. Data Mining and Knowledge Discovery, 32:1179–1199, 2018.
- G. Widmer and M. Kubat. Learning in the presence of concept drift and hidden contexts. Machine learning, 23(1):69–101, 1996.
- Nina Wiedemann and Martin Raubal. Traffic forecasting on traffic moving snippets. arXiv preprint arXiv:2110.14383, 2021. URL <https://arxiv.org/abs/2110.14383>.
- Christian Schroeder de Witt, Catherine Tong, Valentina Zantedeschi, Daniele De Martini, Alfredo Kalaitzis, Matthew Chantry, Duncan Watson-Parris, and Piotr Bilinski. RainBench: Towards Data-Driven Global Precipitation Forecasting from Satellite Imagery. Proceedings of the AAAI Conference on Artificial Intelligence, 35(17):14902–14910, May 2021. ISSN 2374-3468. URL <https://ojs.aaai.org/index.php/AAAI/article/view/17749>. Number: 17.
- M. K. Wouter. An introduction to domain adaptation and transfer learning. arXiv preprint arXiv:1812.11806, abs/1812.11806, 2018. URL <https://arxiv.org/abs/1812.11806>.
- Tianfan Xue, Jiajun Wu, Katherine Bouman, and Bill Freeman. Visual dynamics: Probabilistic future frame synthesis via cross convolutional networks. In Advances in Neural Information Processing Systems, pages 91–99, 2016.
- Lei Zhang, Xun Wang, Nicholas Penwarden, and Qiang Ji. An image segmentation framework based on patch segmentation fusion. In 18th International Conference on Pattern Recognition (ICPR’06), volume 2, pages 187–190. Ieee, 2006.
- Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In Deep learning in medical image analysis and multimodal learning for clinical decision support, pages 3–11. Springer, 2018.
- Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Transactions on Medical Imaging, 2019.