WEATHER4CAST

Solving the Weather4cast Challenge via Visual Transformers for 3D Images

Yury Belousov,[1] Sergey Polezhaev,[2] Brian Pulfer[1]
team “team-name”

December 8, 2022
36th Conference on Neural Information Processing Systems

[1]University of Geneva, Switzerland

[2]Neiro AI, USA

Weather4cast challenge

- Challenge proposed by the *Institute of Advanced Research in Artificial Intelligence* (IARAI)
- The goal of the challenge is to predict the rainfall events in the following 8-hours given a 1-hour context
- Predictions are made on a small spatial crop of the input but with a higher resolution
- Data is provided for years 2019 and 2020 from different regions around the world

Model input

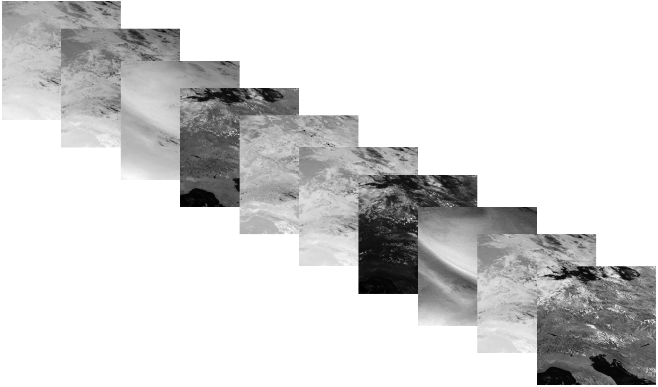

Figure 1: Example of satellite images for region *boxi_0034* in *2019*.

Shape of an input to a model — $(11, 4, 252, 252)$:

- 11 is the number of bands spectral satellite images
- 4 is the time dimension (1 preceding hour $\times$ 4 step, i.e. evenly divided into slots of 15 minutes each)
- 252×252 is the shape of a satellite region.

Model prediction

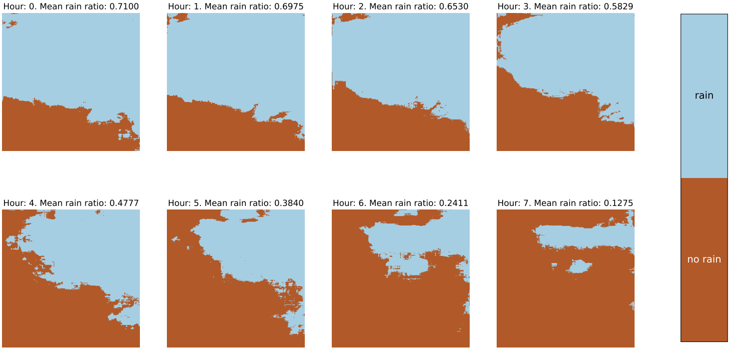

Figure 2: Example of model predictions for region *roxi_0004* in *2020*.

Shape of a prediction — $(32, 252, 252)$:

- 32 is the time dimension (8 next hours $\times$ 4 step with the same time discretization)
- 252×252 is the shape of a rainfall region
- But the spatial resolution of the satellite images is about six times lower than the resolution of the ground radar.

Metric

Performances are measured as the Intersection over Union of the predicted rainfall events $\mathcal{P}$ and the ground truth $\mathcal{G}$:

$$\mathsf{IoU}(\mathcal{P}, \mathcal{G}) = \frac{|\mathcal{P} \cap \mathcal{G}|}{|\mathcal{P} \cup \mathcal{G}|}$$

VIVIT[1]

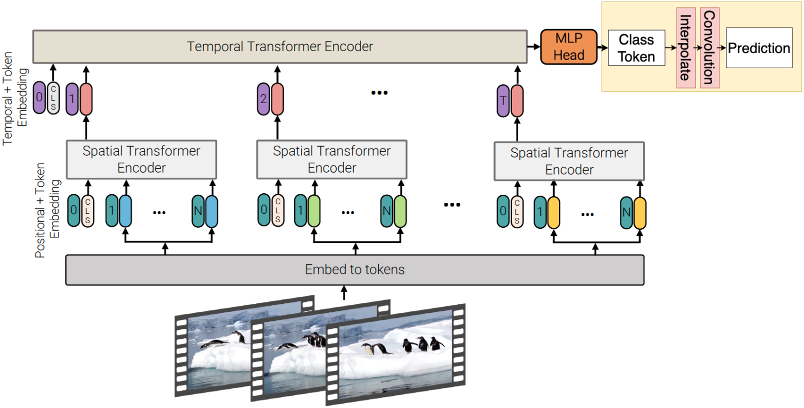

Figure 3: Our VIVIT architecture adaptation

[1] Anurag Arnab et al. "Vivit: A video vision transformer". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 6836–6846.

SWIN-UNETR[2]

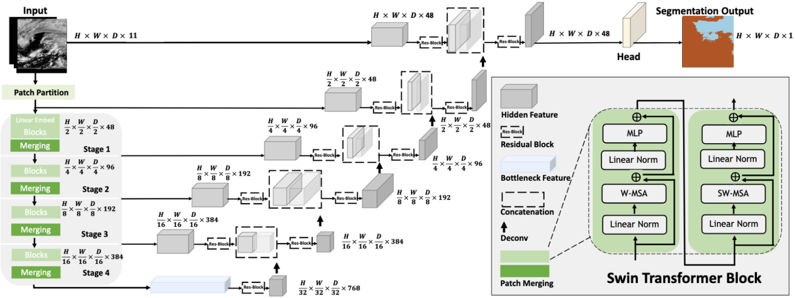

Figure 4: Swin-UNETR architecture

[2] Ali Hatamizadeh et al. "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images". In: *International MICCAI Brainlesion Workshop*. Springer. 2022, pp. 272–284.

Various input transformations for SWIN-UNETR

Submission name	Total mean	2019 mean	2020 mean
Repeat-interleave Epoch 3	**0.252**	**0.262**	**0.241**
Channel Convolution Epoch 1	0.244	0.258	0.230
Upsample Epoch 1	0.224	0.256	0.192

All versions were trained for minimum 4 epochs

16-bit training & gradient checkpointing

Submission name	Total mean	2019 mean	2020 mean
32-bit training Epoch 3	**0.252**	**0.262**	0.241
16-bit training Epoch 3	**0.252**	0.253	**0.250**

Almost identical results with or w/o 16-bit training

Baseline improvements

- an attention grid[3]
- changing the activation from RELU to RRELU
- changing normalization from batch to instance
- replacing transpose convolution with upsampling and regular convolution.

Submission name	Total mean	2019 mean	2020 mean
base	0.213	0.243	0.183
improved	0.245	**0.274**	0.217
improved & w/o convtranspose	**0.246**	0.267	**0.225**

[3] Ozan Oktay et al. "Attention u-net: Learning where to look for the pancreas". In: *arXiv preprint arXiv:1804.03999* (2018).

Model-independent configurations: Loss

Model type	Loss	Total mean	2019 mean	2020 mean
BASELINE	IoU	0.190	0.210	0.171
	bce	0.213	0.243	0.183
SWIN-UNETR	IoU	0.190	0.206	0.174
	dice focal	0.210	0.228	0.192
	bce	**0.252**	**0.262**	**0.241**

Model-independent configurations: Dataset

A discrepancy between the training and validation datasets[4]:

- $\mu = 2.53 \times 10^{-2}$, $\max = 6.78 \times 10^{-2}$ for training
- $\mu = 4.79 \times 10^{-2}$, $\max = 11.3 \times 10^{-2}$ for validation

Model type	Submission name	Total mean	2019 mean	2020 mean
BASELINE	train. Epoch 23	0.213	0.243	0.183
	train & val. Epoch 24	**0.222**	**0.252**	**0.192**
	train & val. Epoch 53	0.166	0.185	0.147

[4] for *roxi_0007* in 2020

Model-independent configurations: Threshold

Model type	Submission name	Total mean	2019 mean	2020 mean
SWIN-UNETR	0.5 threshold	**0.252**	**0.262**	**0.241**
	0.2 threshold	0.227	0.248	0.207
	0.65 threshold	0.194	0.204	0.183

Majority voting

- Generate predictions of different models
- The most frequent option determines the final prediction for each pixel
- If most models predict it will rain at a given place at a given moment, that will be the final prediction and vice-versa.

Submission name	Total mean	2019 mean	2020 mean
Best individual model	0.252	0.262	0.241
Majority voting	**0.265**	**0.289**	**0.242**

This approach could be further optimized by excluding worst models

What we haven't had time to test

- **Optimizer:** changing from AdamW to AdaBelief[5]
- **Temporal shift:** predict the time deltas starting from the second time step: $t'_0 = t_0, t'_i = t'_{i-1} + t_i$ for $i \geq 1$, where t_i is a raw model's delta prediction from time $i-1$ to i and t'_i is the final prediction
- **Embedding:** either for a region or time (year/season/month)
- **Masking:** proper masking missing measurements

[5] Juntang Zhuang et al. "Adabelief optimizer: Adapting stepsizes by the belief in observed gradients". In: *Advances in neural information processing systems* 33 (2020), pp. 18795–18806.

Results: Heldout

Submission name	Total mean	2019 mean	2020 mean
Official BASELINE	*0.255*	*0.259*	*0.251*
BASELINE bce improved. Epoch 15	0.270	0.261	0.278
SWIN-UNETR bce. Epoch 3	0.281	0.283	0.280
Majority vote	**0.300**	**0.296**	**0.303**
Take best prediction per region	0.302	0.301	0.303

Conclusions

- Our work to tackle the Weather4Cast competition:
    - Model-independent configurations
    - Baseline improvements
    - Vivit model adaptation
    - SWIN-UNETR model adaptation
- Ensembling yields the most competitive results
- We are placed 3^{rd} ex-aequo.

Thanks! Questions?

Code:

https://github.com/bruce-willis/weather4cast-2022

Paper:

https://arxiv.org/abs/2212.02456

References i

Arnab, Anurag et al. “Vivit: A video vision transformer”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 6836–6846.

Hatamizadeh, Ali et al. “Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images”. In: *International MICCAI Brainlesion Workshop*. Springer. 2022, pp. 272–284.

Oktay, Ozan et al. “Attention u-net: Learning where to look for the pancreas”. In: *arXiv preprint arXiv:1804.03999* (2018).

Zhuang, Juntang et al. “Adabelief optimizer: Adapting stepsizes by the belief in observed gradients”. In: *Advances in neural information processing systems* 33 (2020), pp. 18795–18806.