Retrieval augmented diffusion models for time-series forecasting

Neural Networks 2024/25
21/07/2025

Valerio Baldi 1940729
Saverio Dieni 1946039

Task:
time-series forecasting

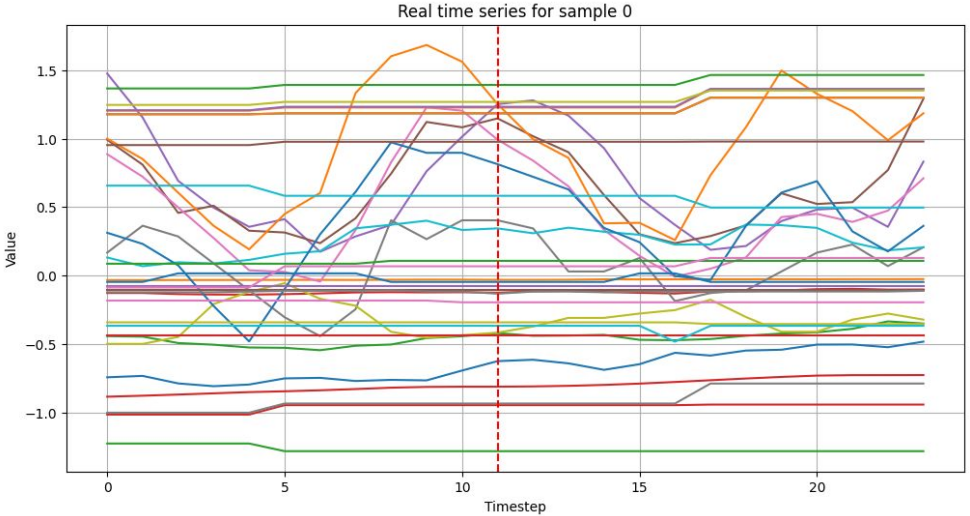
Minimization problem: find the probability distribution p parameterized by θ of having x^P given x^H that best fits the the real probability distribution

Conditional Time Series Diffusion Models

Idea: in the forward process you can inject some controlled noise into the signal, then in the backward process you learn how to reconstruct the original signal.

Dataset:

House price dataset HouseTS from Kaggle, we considered 34 features and took 24 timesteps series (each step of 1 month) where 12 steps were given to forecast the last 12.

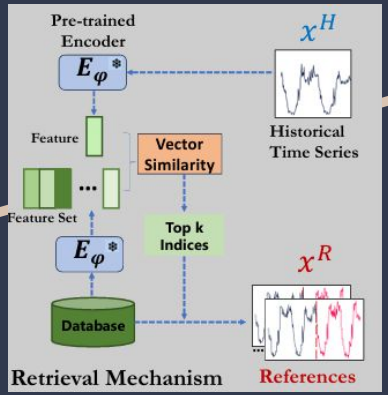


Retrieval database

There are 2 main problems in time-series generation:lack of a meaningful guidance

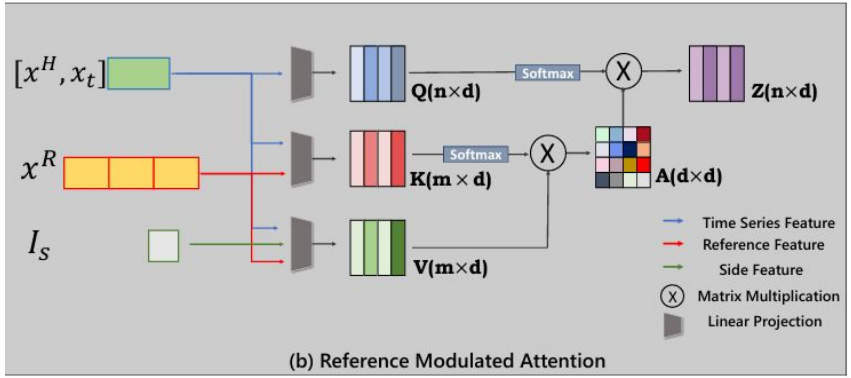
- insufficient size
- unbalance of the dataset

We embed the head of each time-series using a pre-trained autoencoder model, we build the retrieval database on the embeddings indexed with FAISS. Now we can find top k nearest neighbours based on those embeddings and take the tails of those series as a reference for the generation process.

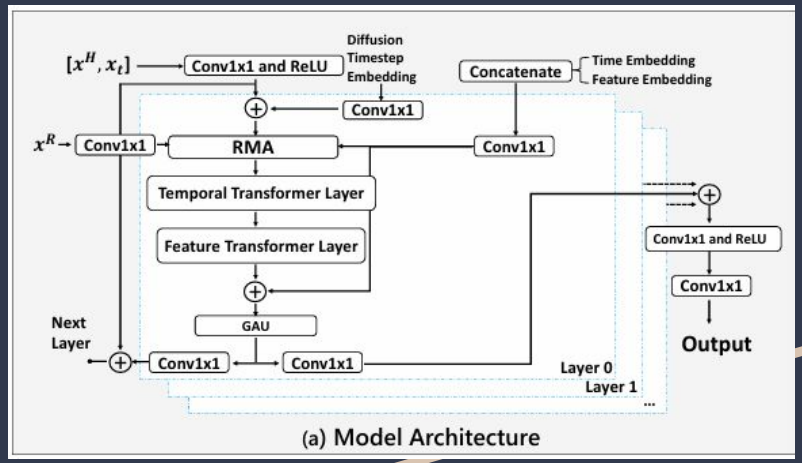


Reference Moduled Attention (RMA)

The main novelty of the model presented in the paper is the Reference Moduled Attention (RMA) which uses the time-series references to guide the denoising process. It is an alternative to the Cross-Attention Module and it is specifically designed to exploit the references and the side information.



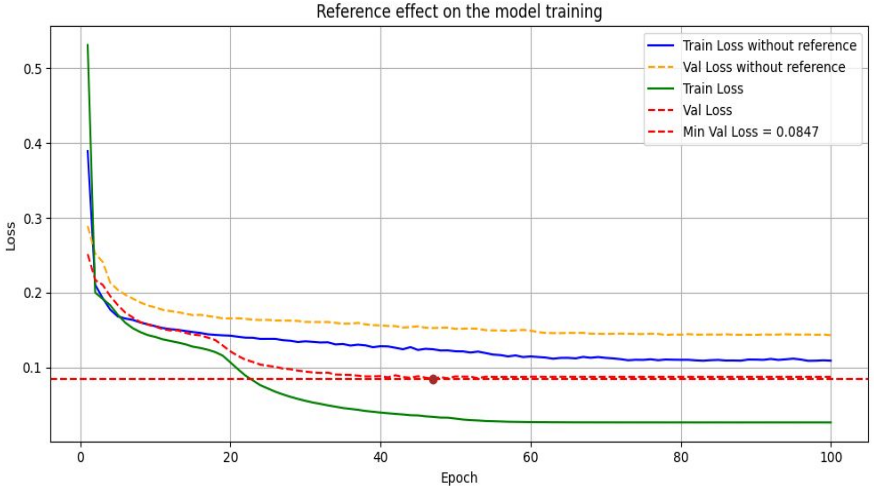
Retrieval Augmented Time series Diffusion model (RATD)



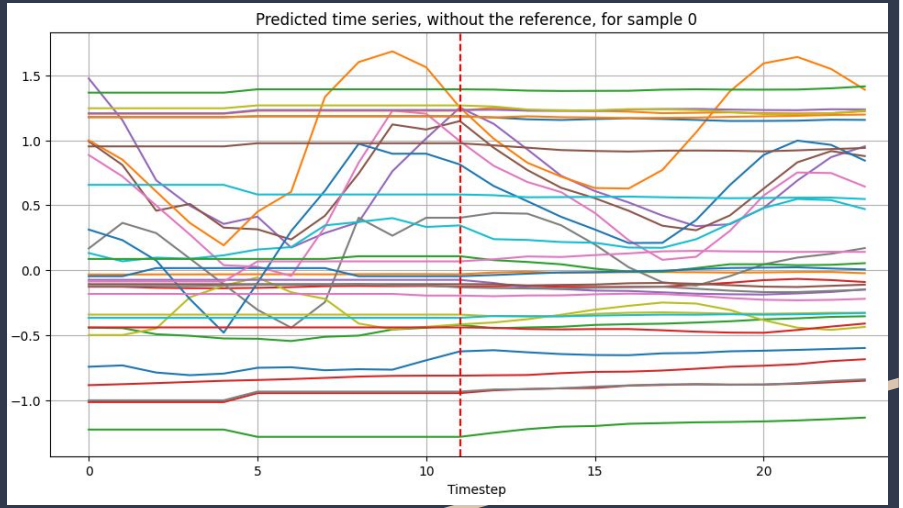
(a) Model Architecture

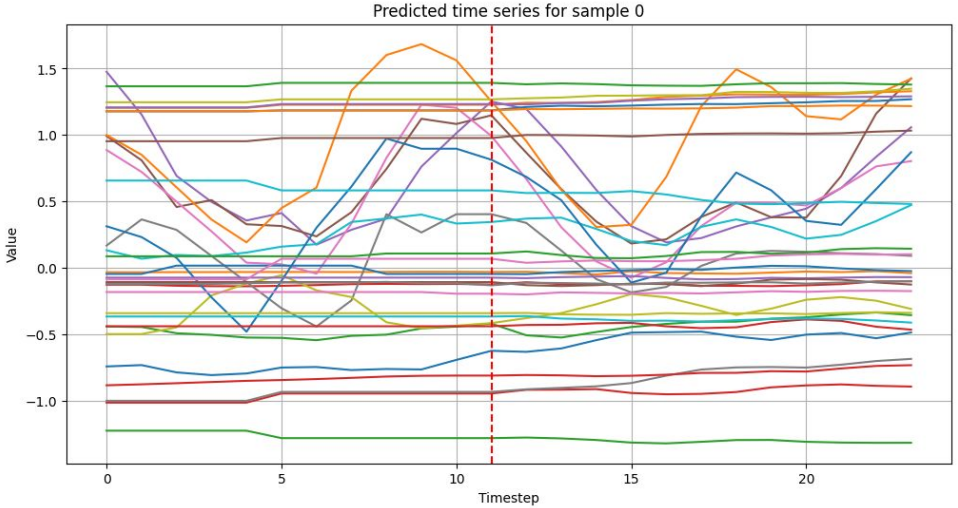
Experiment

We trained our RATD model for 100 epochs and reached a MSE on the validation dataset of 0.0847.



Results


Predicted time series, without the reference, for sample 0
1.5
1.0
0.5
0.0
−0.5
−1.0
0
5
10
15
20
Timestep


Predicted time series for sample 0
Value
1.5
1.0
0.5
0.0
−0.5
−1.0
0
5
10
15
20
Timestep

Multimodal RATD

We expanded our RATD model by exploiting satellite images in the HouseTS dataset related to the time-series.
We used the FiLM (Feature-wise Linear Modulation) technique to integrate the images in the model.

$$\text{FiLM}(x \mid \gamma_x, \beta_x) = \gamma_x \odot x + \beta_x$$
$$\text{FiLM}(xr \mid \gamma_{xr}, \beta_{xr}) = \gamma_{xr} \odot xr + \beta_{xr}$$

where γ, β are two modulation parameter computed from an embedding y:

$$\text{Concat}(\gamma, \beta) = f_\theta(y)$$

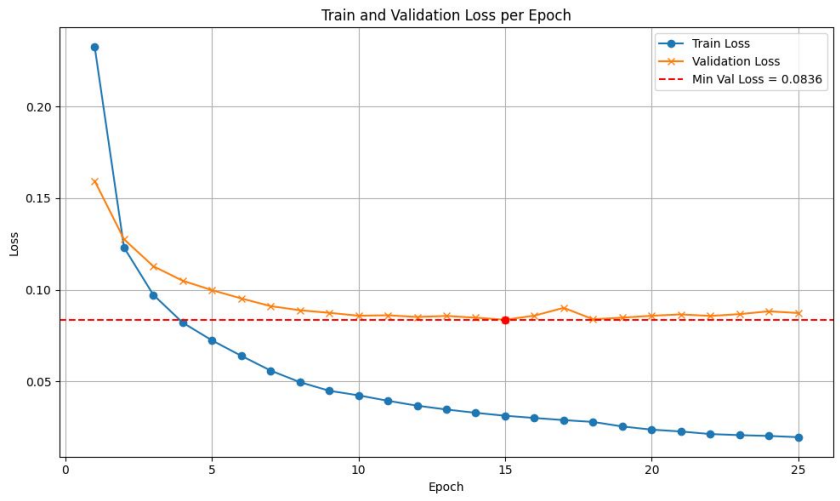
CLIP encoder

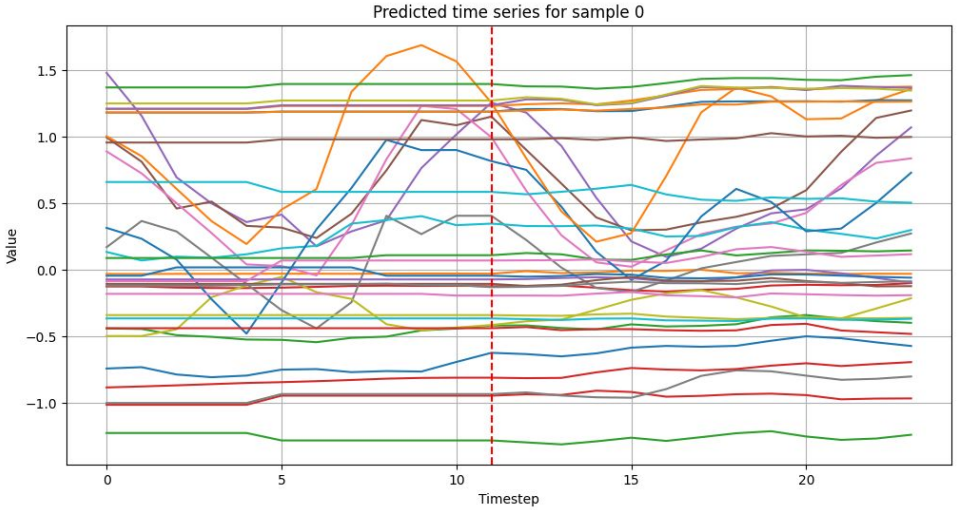
To embed the images we used an encoder based on the pre-trained CLIP encoder, then we appropriately reshaped y to obtain the parameters γ, β.

$$y = \text{CLIPEnc}(\text{img}_x)$$

$$f_\theta(y) = \text{MLP}(\text{Conv1D}(\text{LN}(y)))$$

Results


Train and Validation Loss per Epoch
Train Loss
Validation Loss
Min Val Loss = 0.0836
Loss
Epoch
0.20
0.15
0.10
0.05
0
5
10
15
20
25


Predicted time series for sample 0
Value
Timestep
1.5
1.0
0.5
0.0
−0.5
−1.0
0
5
10
15
20