Filter Banks:

from Low-Rank Representations to Deep Models for Efficient Time Series Forecasting

Ashutosh Vaishnav[1], Mohsen Amidzadeh[1], Teemu Kämäräinen[2], Matti Siekkinen[1], Mario Di Francesco[1]

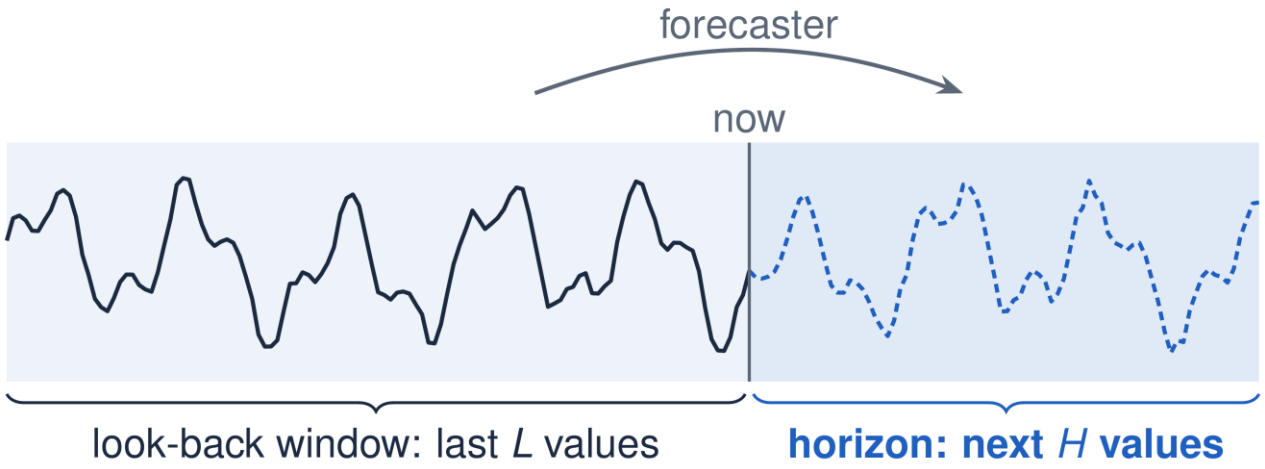
[1]Aalto University [2]University of Helsinki

NeurIPS 2026



Simple linear models can rival large transformers

The task: long-term forecasting



Standard benchmarks

Dataset	Channels	Time steps
ETTh1, ETTh2	7	17,420
ETTm1, ETTm2	7	69,680
Weather	21	52,696
Electricity	321	26,304
Traffic	862	17,544

Model families tried on them

Linear
DLinear,
FITS, SparseTSF

Transformer
PatchTST,
iTransformer

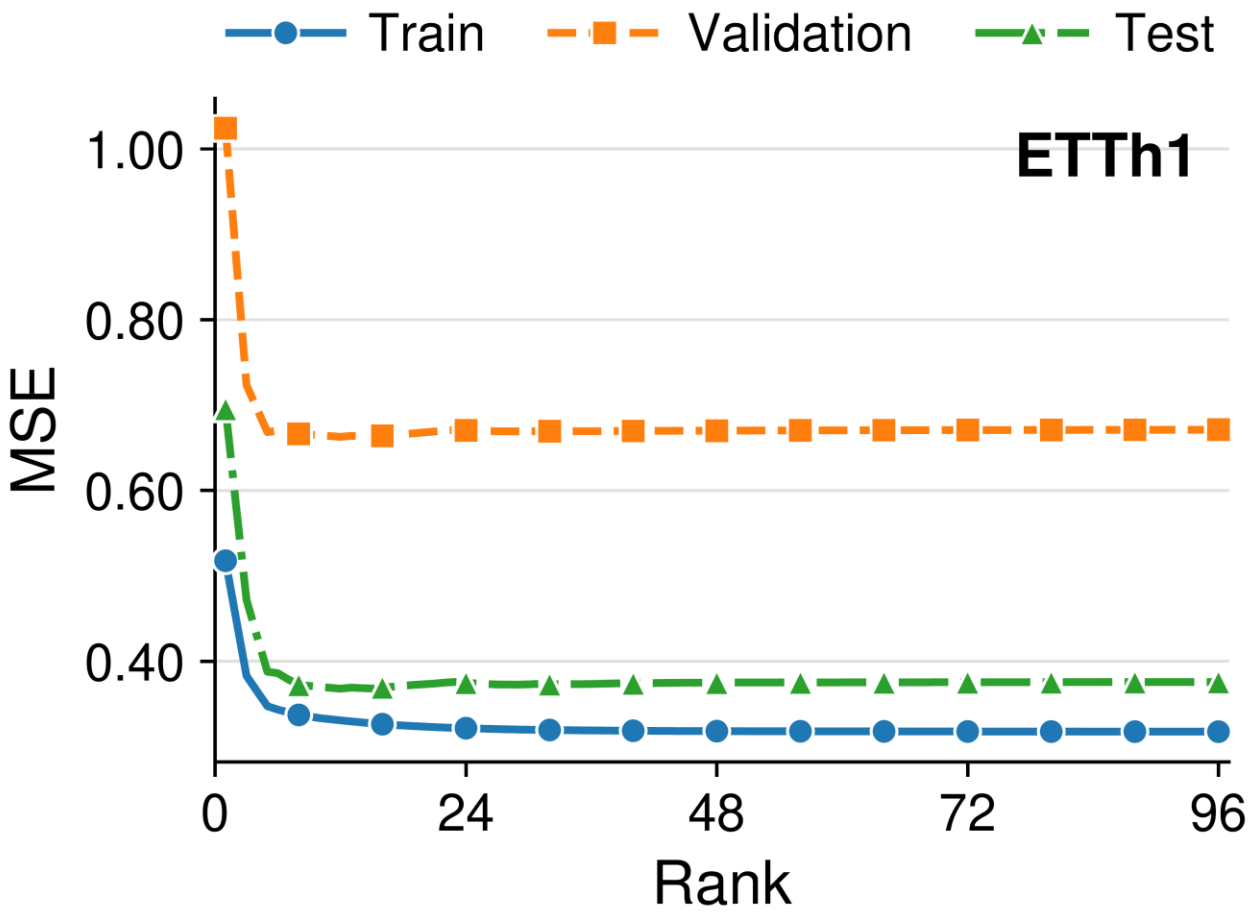
MLP mixer
TimeMixer

KAN
TimeKAN

State-space
TimePro

How complex are these forecasting tasks really?

Long-term forecasting tasks are intrinsically low-rank



24.7x

average compression of the linear model

rank 3–24

is optimal, out of a full rank of 96–720

about the same test MSE

as full-rank least squares
(0.64% better on average)

Reduced-rank regression, look-back $L = 720$, ETT and Weather datasets, 20 dataset-horizon settings.

Kernel analysis shows: the forecaster is a small set of horizon-invariant filters

horizon-varying coefficients

$$\hat{y}_{i,n} = \sum_{l=1}^{L} \sum_{k=1}^{n_f} \gamma_{k,n} \, f_{k,l} \, x_{i,L-l+1}$$

horizon-invariant filters

$$n_f = r$$

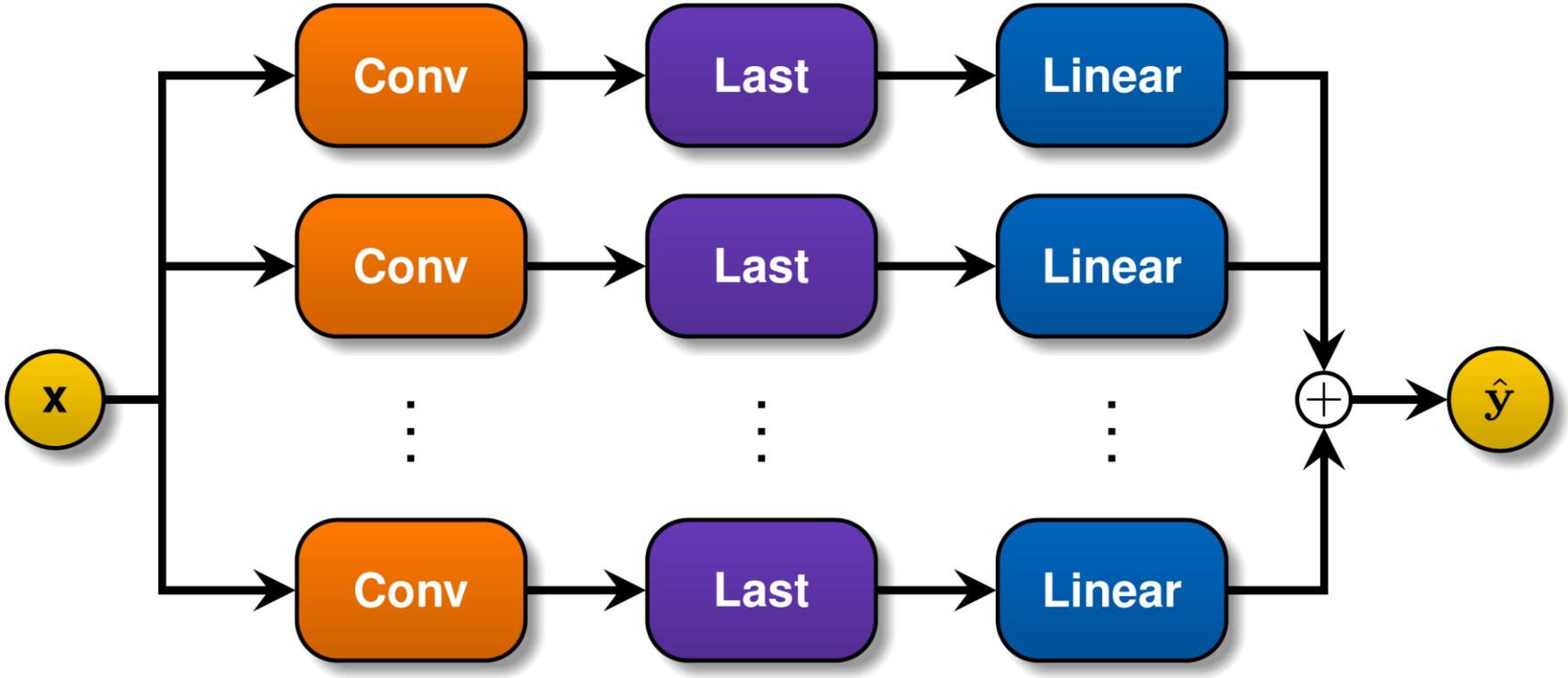
number of filters rank

Parameters: $L \cdot H \rightarrow n_f(L + H)$

518,400 → 14,400

for $L = H = 720$ at rank 10 (97% fewer)

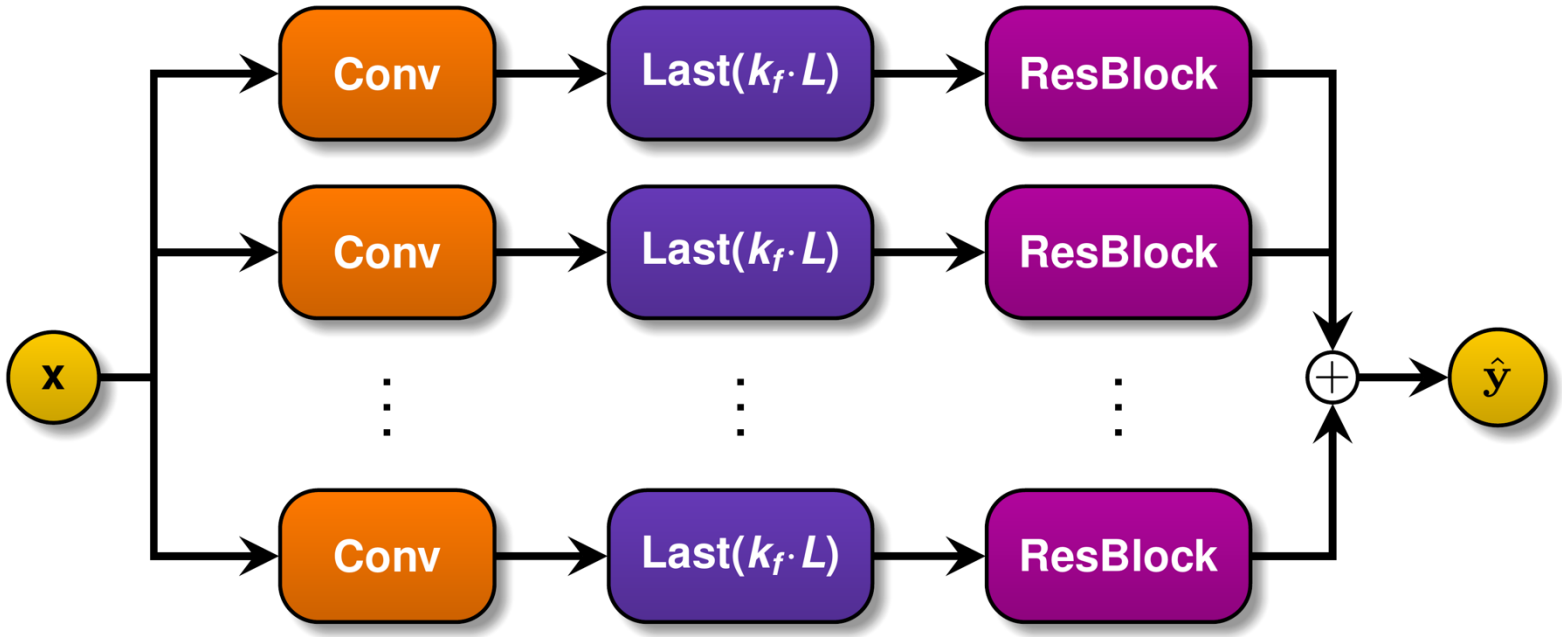
FB-LR turns this structure into a learnable architecture



- n_f convolution filters
- keep only the last output sample
- linear weights per branch, then sum

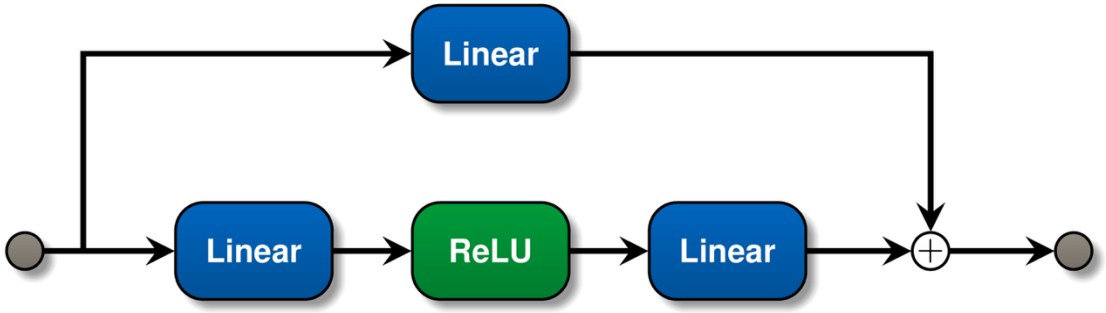
rank ≤ n_f, learned end to end, no kernel to choose

FB-deep relaxes the rank constraint with depth



keep a fraction k_f of the outputs per branch

process with a residual block



$$\mathrm{rank}(\mathbf{W}) \leq \min\{n_f k_f L,\ H,\ L\}$$

$k_f = 1/L$ recovers FB-LR

Competitive accuracy at a fraction of the computation cost

Accuracy

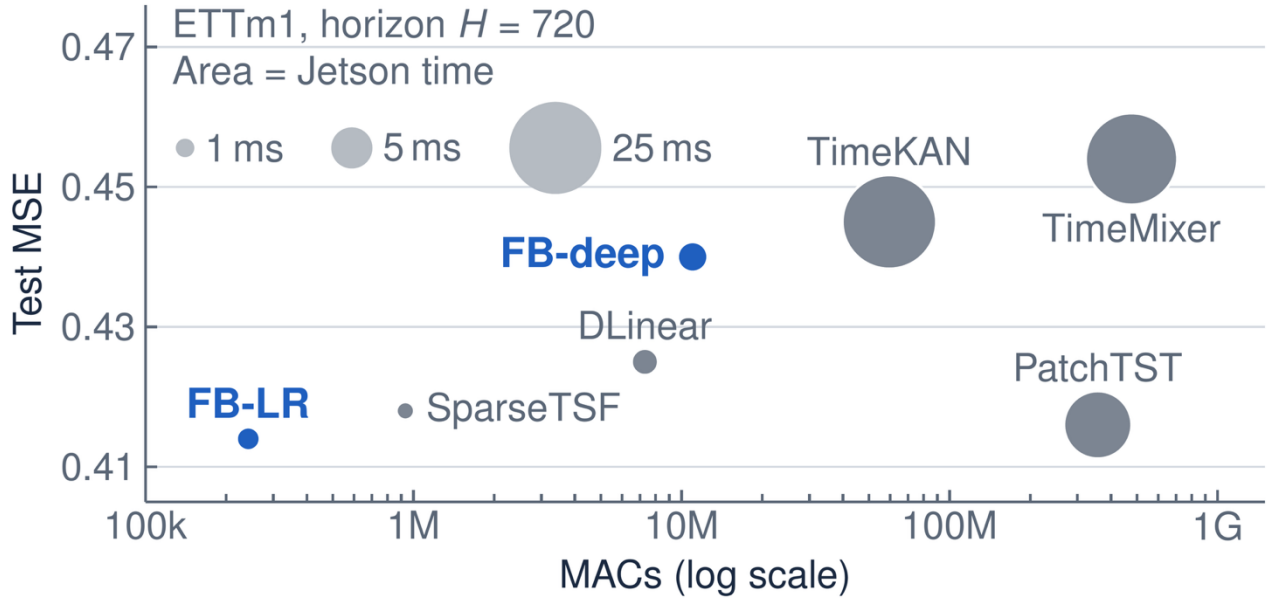
22 / 28

settings where **FB-LR** beats the full-rank DLinear

11 / 12

settings where **FB-deep** has the lowest MSE on Weather, Electricity and Traffic (7 baselines compared)

Cost on ETTm1, *L* = 720



32.5x

fewer MACs required by FB-deep than PatchTST, ~5x faster on a Jetson Orin AGX

Low rank is a design principle, not only a compression trick

1. Popular long-term forecasting benchmarks are intrinsically low-rank
2. Filter banks expose rank as a tunable, interpretable hyperparameter
3. They are competitive with the state of the art at far lower cost, also on edge hardware



Code

github.com/ashutoshvai1/FilterBanks