

Multi-Modal View Enhanced Large Vision Models for Long-Term Time Series Forecasting

多模态视图增强的大视觉模型在长期时间序列预测中的应用

2025/11/23

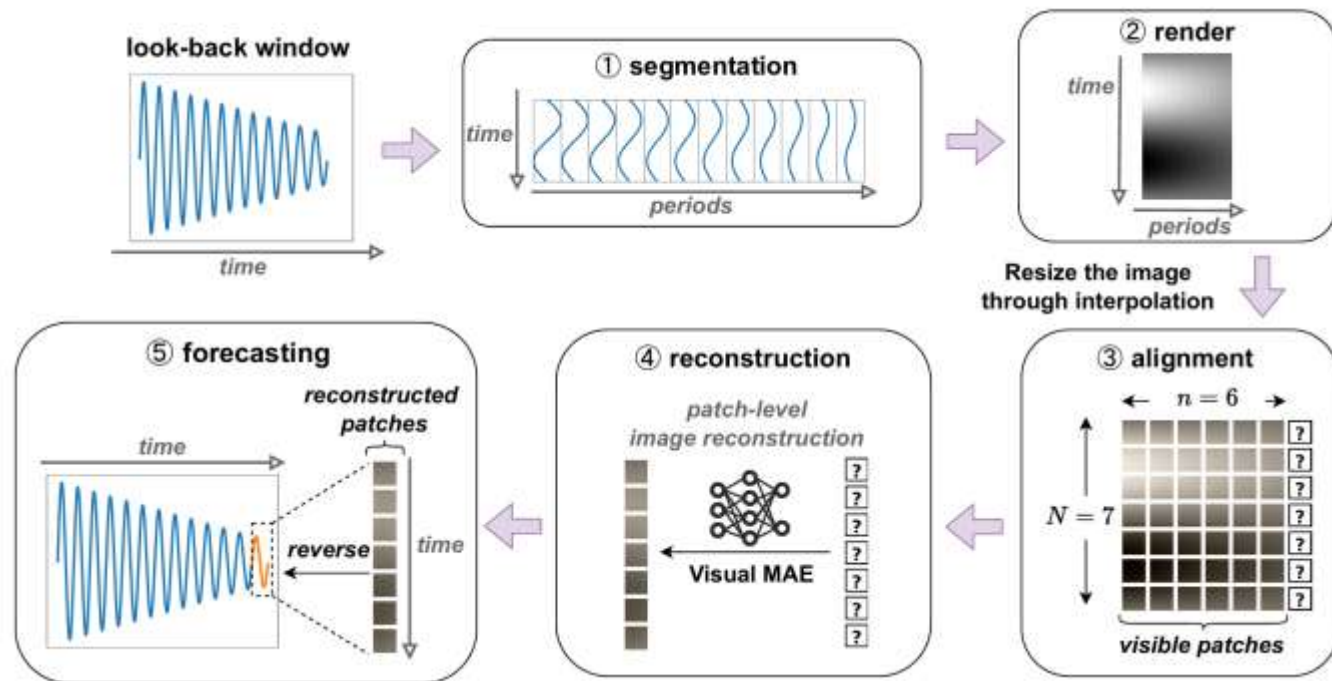
I Introduction

I Introduction

研究背景

长期时间序列预测的应用覆盖能源、医疗、交通、智慧城市等关键领域，需精准预测未来长时步数据。如今时间序列可转化为数值、图像、文本三种模态，其中图像模态适配大视觉模型，能挖掘周期性等隐藏模式。

大视觉模型如 ViT、BEiT，因预训练于大规模图像数据，且图像通道的连续像素序列与单变量时间序列结构相似，其比处理离散 tokens 的 LLM 更适配时间序列；典型代表为 VisionTS视觉模型



VisionTS视觉模型架构



I Introduction

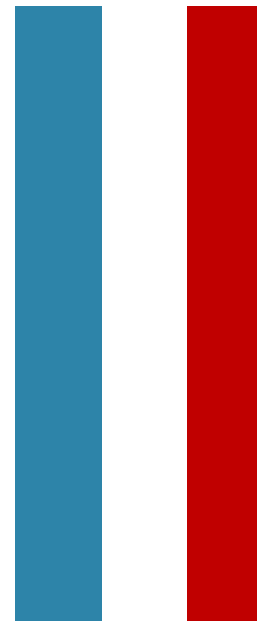
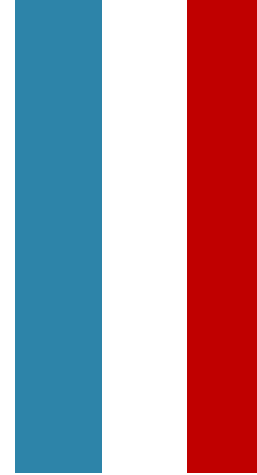
当前研究不足

- 现有 LVM-based 预测存在预测周期归纳偏置，过度关注周期性而忽略全局趋势。
- 融合策略简单，采用通用的嵌入层融合方式，未考虑不同模态的固有特性
- Time-VLM 等多模态模型融合策略不合理、文本模态增益有限且计算开销大，需设计更高效的多模态融合框架

研究目标

- 设计新的分解式多模态视图框架DMMV，旨在通过分解与多视图协同的设计，让 LVM 专注发挥其周期性建模优势，同时通过专门的趋势建模模块弥补其缺陷，从根本上缓解归纳偏置带来的负面影响。从而构建高效且适配时间序列特性的多模态融合机制。
- 设计 DMMV 框架时排除增益有限的文本模态，仅微调 LVM 的归一化层在保证性能的同时降低模型复杂度与推理成本，开发出兼具高性能与实用性的 LTSF 模型

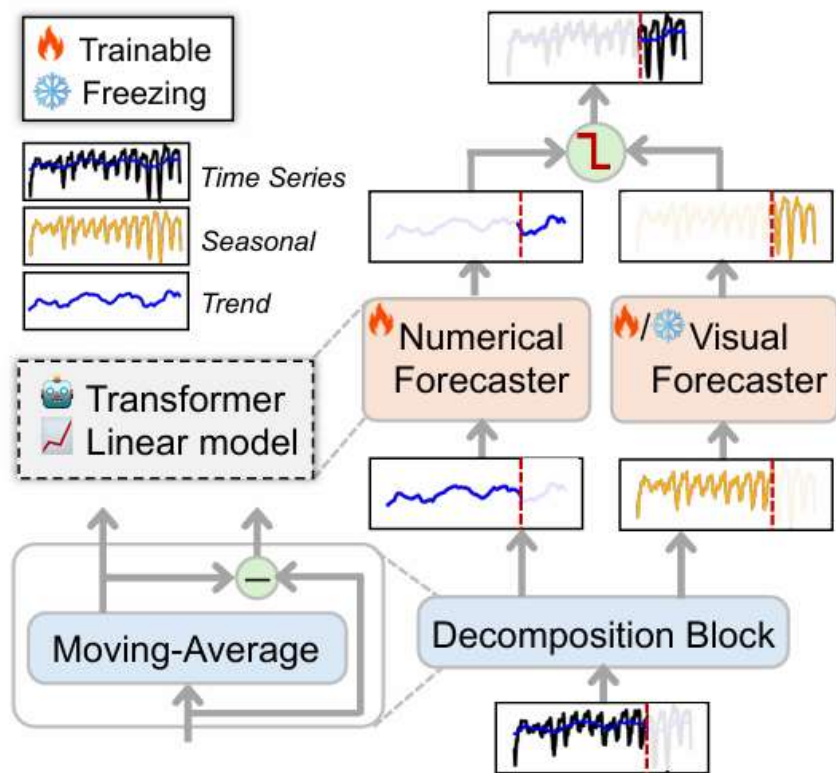
II Methods



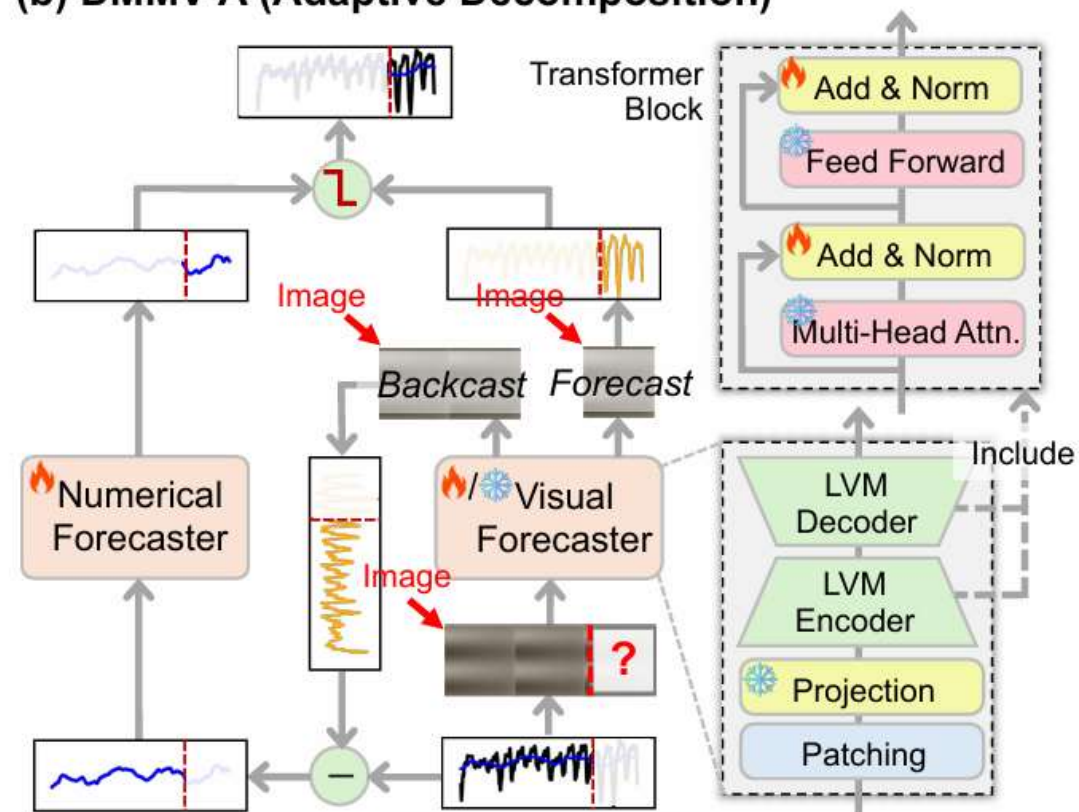
II Methods

总体框架

(a) DMMV-S (Simple Decomposition)



(b) DMMV-A (Adaptive Decomposition)



II Methods

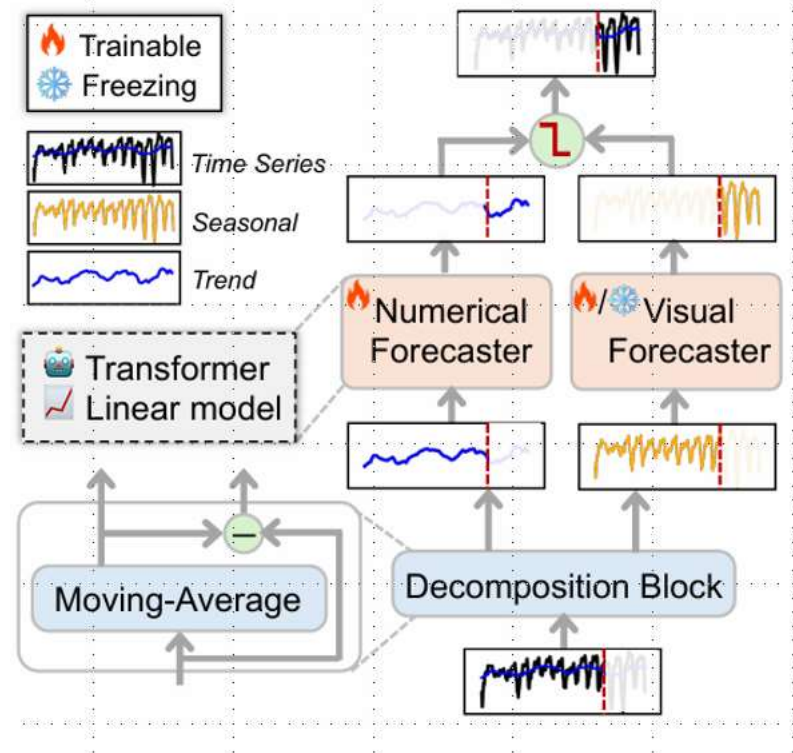
DMMV-S基于简单分解的实现

DMMV-S采用了一种经典的时间序列分解方法——移动平均。对于一个输入的多元时间序列 X ，它首先采用通道的策略，即分别处理每一个维度的单变量时间序列

分解过程通过一个滑动窗口来获取平滑数据，从而提取出低频的趋势部分，而原始序列减去趋势部分后，剩下的高频波动则被视为季节性部分。这个过程可以用以下公式表示

$$\mathbf{x}_{\text{trend}}^i = \text{Moving-Average}(\text{Padding}(\mathbf{x}^i)), \quad \mathbf{x}_{\text{season}}^i = \mathbf{x}^i - \mathbf{x}_{\text{trend}}^i, \quad 1 \leq i \leq D$$

(a) DMMV-S (Simple Decomposition)



II Methods

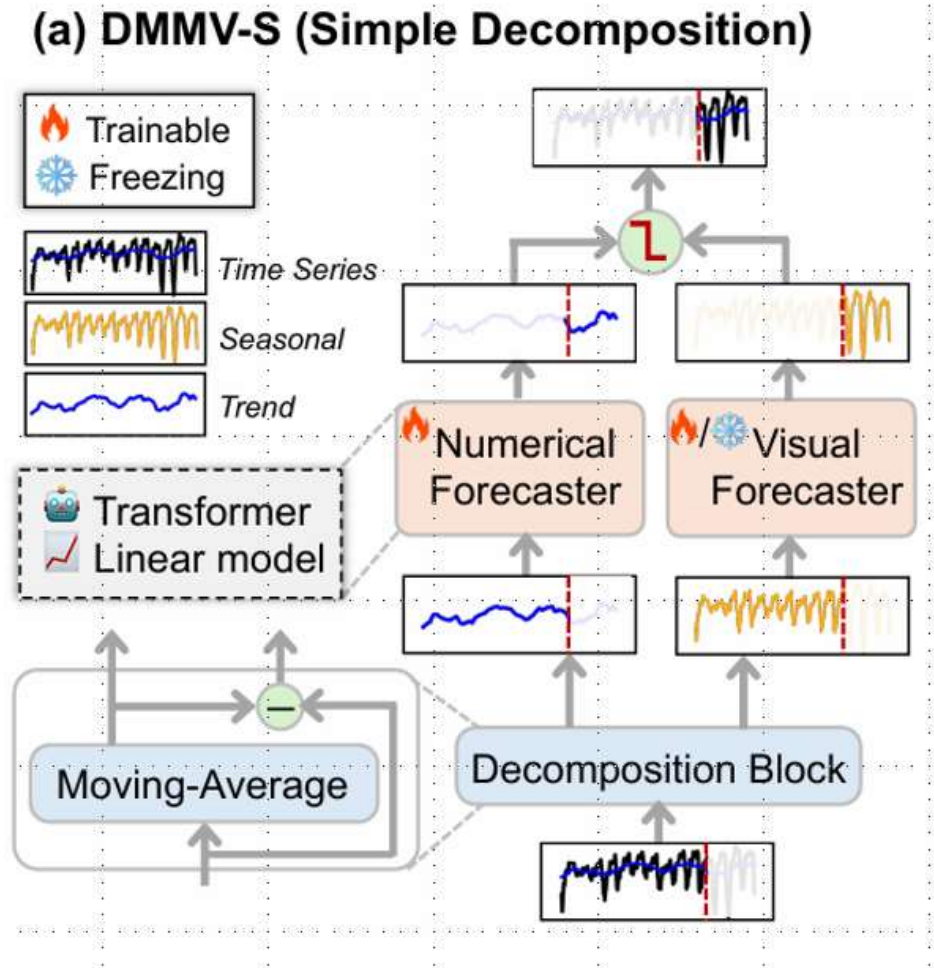
DMMV-S基于简单分解的实现

分解之后，这两个部分被送往不同的专家预测器：

数值预测器：负责处理趋势部分。由于趋势通常变化平缓，具有较强的线性特征，采用了一个简单的线性模型进行预测

视觉预测器：负责处理季节性部分。它首先将这个一维的季节性序列，通过一种名为周期性成像的技术转换成二维图像。具体来说，它根据序列的主要周期将其切分成多段，然后堆叠成一个矩阵。接着，这个图像被送入一个预训练好的LVM。LVM凭借其强大的图像模式识别能力，能够精准捕捉季节性序列中的周期性规律，并预测未来的季节性变化。

最后，两个预测器分别得到的趋势预测值和季节性预测值，通过一个可学习的门控机制进行加权融合，得到最终的预测结果。



II Methods

DMMV-A: 基于自适应分解的实现

与DMMV-S的“先分解，后预测”不同，DMMV-A巧妙地利用了视觉预测器本身的特性，来自动学习如何分解序列。它的核心是一种回溯-残差机制

DMMV-A不再对输入序列进行预先分解。而是直接将原始的序列图像化后送入视觉预测器。但此时，模型的任务不是预测未来，而是“回溯”历史，即重建输入序列自身。为了高效地完成这个任务，论文设计了一种名为BCMASK的掩码策略。

BCMASK分两步完成回溯：首先遮住图像的左半部分，用右半部分来预测左半部分；然后反过来，遮住右半部分，用左半部分来预测它。通过拼接这两次重建的结果，就得到了一个完整的对输入图像的回溯

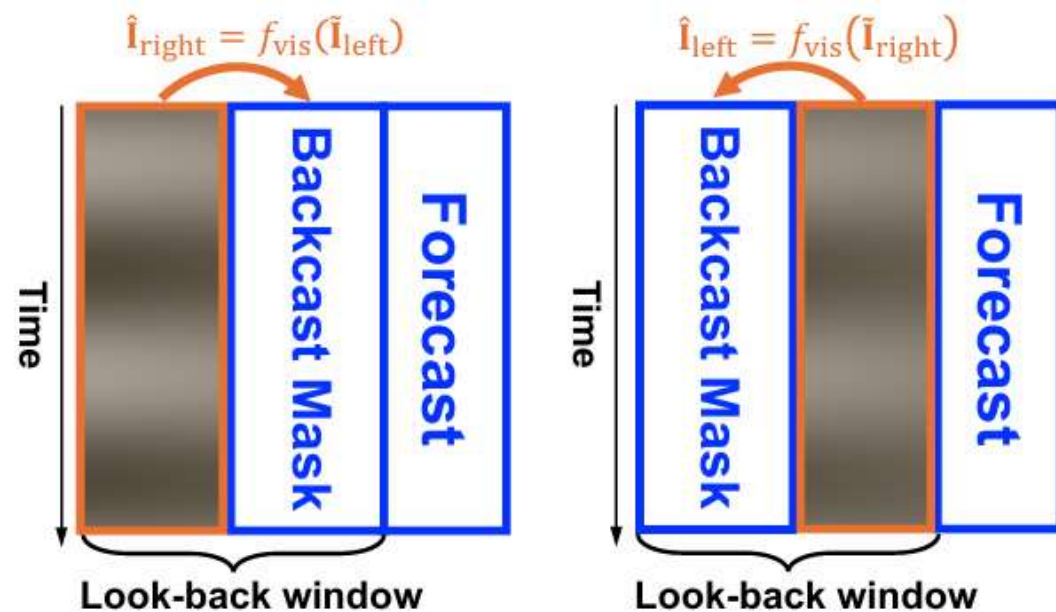


Figure 4: An illustration of BCMASK.

II Methods

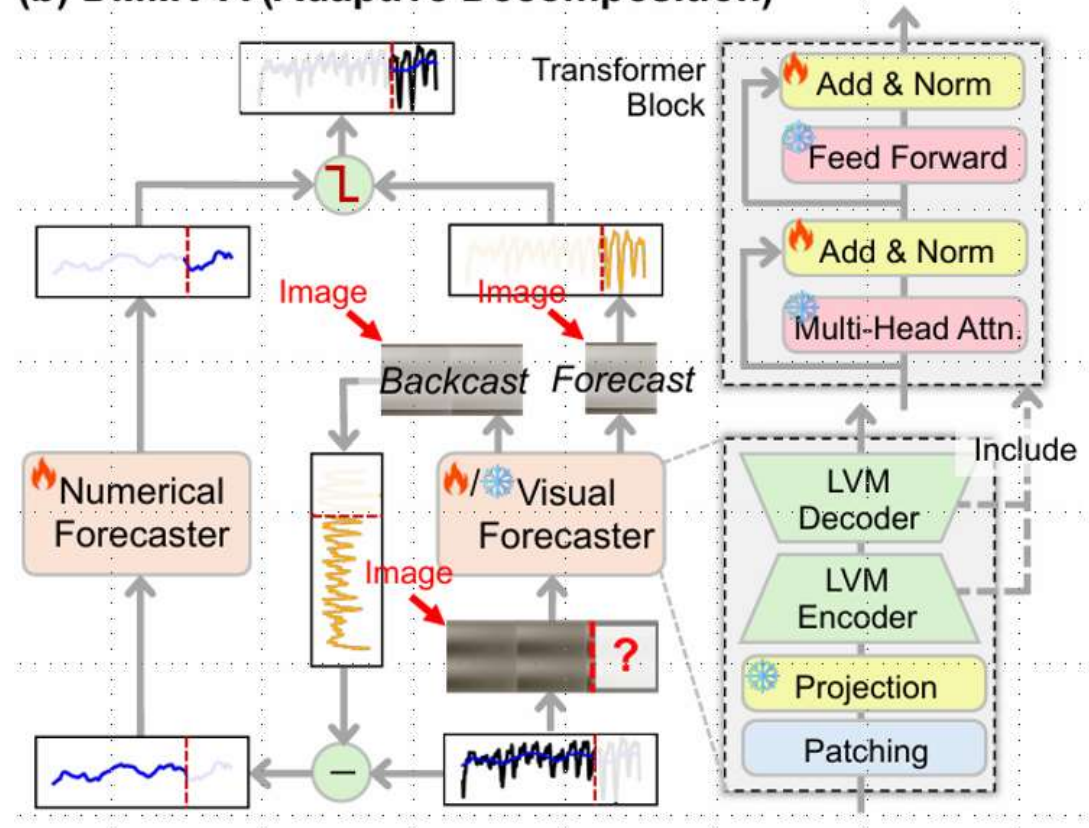
DMMV-A: 基于自适应分解的实现

由于视觉预测器 存在周期性归纳偏置，它重建出的回溯序列会天然地保留原始序列中最主要的周期性成分，而过滤掉那些非周期的趋势信息。因此，可以被看作是模型自动提取出的季节性部分。而原始序列与这个回溯序列的残差，则恰好代表了被剥离周期性后的趋势部分

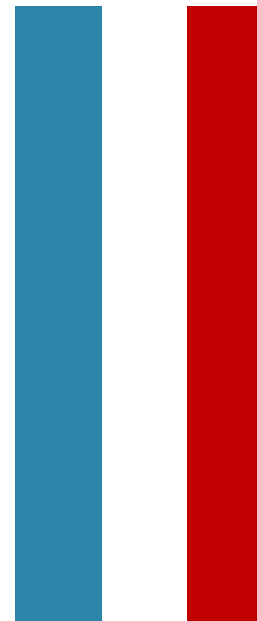
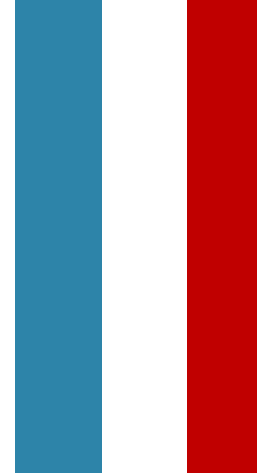
$$\Delta \mathbf{x}^i = \mathbf{x}^i - \hat{\mathbf{x}}^i$$

分解完成后，流程就和DMMV-S类似了。趋势部分被送入数值预测器。而视觉预测器 则在送入原始图像化序列的同时，直接进行未来预测，得到季节性预测

(b) DMMV-A (Adaptive Decomposition)



III Experiment



III Experiment

数据集:

数据集	变量数D	序列长度	采样频率	预测步长H	划分比例 (训 / 验 / 测)
ETTh1/2	7	17420	小时	96,192,336,720	60%/20%/20%
ETTm1/2	7	69680	15 分钟	96,192,336,720	60%/20%/20%
Weather	21	52696	10 分钟	96,192,336,720	70%/10%/20%
Illness	321	966	周	24,36,48,60	70%/10%/20%
Traffic	862	17544	小时	96,192,336,720	70%/10%/20%
Electricity	37	26304	小时	96,192,336,720	70%/10%/20%

评估指标: 均方误差 MSE, 平均绝对误差 MAE

III Experiment

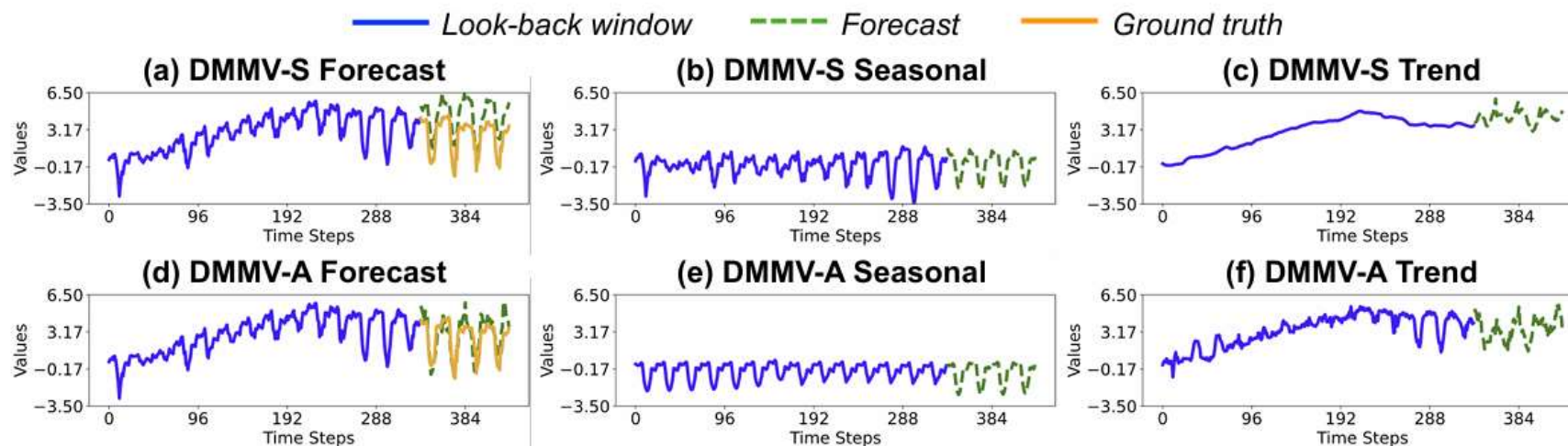
实验结果:

View		Multi-Modal				Visual		Language				Numerical									
Model		DMMV-A		Time-VLM		VisionTS		GPT4TS		Time-LLM		PatchTST		CycleNet		TimesNet		DLinear		FEDformer	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.354	0.389	0.361	0.386	0.355	0.386	0.370	0.389	0.376	0.402	0.370	0.399	0.374	0.396	0.384	0.402	0.375	0.399	0.376	0.419
	192	0.393	0.405	0.397	0.415	0.395	0.407	0.412	0.413	0.407	0.421	0.413	0.421	0.406	0.415	0.436	0.429	0.405	0.416	0.420	0.448
	336	0.387	0.413	0.420	0.421	0.419	0.421	0.448	0.431	0.430	0.438	0.422	0.436	0.431	0.430	0.491	0.469	0.439	0.416	0.459	0.465
	720	0.445	0.450	0.441	0.458	0.458	0.460	0.441	0.449	0.457	0.468	0.447	0.466	0.450	0.464	0.521	0.500	0.472	0.490	0.506	0.507
	Avg.	0.395	0.414	0.405	0.420	0.407	0.419	0.418	0.421	0.418	0.432	0.413	0.431	0.415	0.426	0.458	0.450	0.423	0.430	0.440	0.460
ETTh2	96	0.294	0.349	0.267	0.335	0.288	0.334	0.280	0.335	0.286	0.346	0.274	0.336	0.279	0.341	0.340	0.374	0.289	0.353	0.358	0.397
	192	0.339	0.395	0.326	0.373	0.349	0.380	0.348	0.380	0.361	0.391	0.339	0.379	0.342	0.385	0.402	0.414	0.383	0.418	0.429	0.439
	336	0.322	0.384	0.357	0.406	0.364	0.398	0.380	0.405	0.390	0.414	0.329	0.380	0.371	0.413	0.452	0.452	0.448	0.465	0.496	0.487
	720	0.392	0.425	0.412	0.449	0.403	0.431	0.406	0.436	0.405	0.434	0.379	0.422	0.426	0.451	0.462	0.468	0.605	0.551	0.463	0.474
	Avg.	0.337	0.388	0.341	0.391	0.351	0.386	0.354	0.389	0.361	0.396	0.330	0.379	0.355	0.398	0.414	0.427	0.431	0.447	0.437	0.449
ETTm1	96	0.279	0.329	0.304	0.346	0.284	0.332	0.300	0.340	0.291	0.341	0.290	0.342	0.299	0.348	0.338	0.375	0.299	0.343	0.379	0.419
	192	0.317	0.357	0.332	0.366	0.327	0.362	0.343	0.368	0.341	0.369	0.332	0.369	0.334	0.367	0.374	0.387	0.335	0.365	0.426	0.441
	336	0.351	0.381	0.364	0.383	0.354	0.382	0.376	0.386	0.359	0.379	0.366	0.392	0.368	0.386	0.410	0.411	0.369	0.386	0.445	0.459
	720	0.411	0.415	0.402	0.410	0.411	0.415	0.431	0.416	0.433	0.419	0.416	0.420	0.417	0.414	0.478	0.450	0.425	0.421	0.543	0.490
	Avg.	0.340	0.371	0.351	0.376	0.344	0.373	0.363	0.378	0.356	0.377	0.351	0.381	0.355	0.379	0.400	0.406	0.357	0.379	0.448	0.452
ETTm2	96	0.172	0.260	0.160	0.250	0.174	0.262	0.163	0.249	0.162	0.248	0.165	0.255	0.159	0.247	0.187	0.267	0.167	0.260	0.203	0.287
	192	0.227	0.298	0.215	0.291	0.228	0.297	0.222	0.291	0.235	0.304	0.220	0.292	0.214	0.286	0.249	0.309	0.224	0.303	0.269	0.328
	336	0.272	0.327	0.270	0.325	0.281	0.337	0.273	0.327	0.280	0.329	0.274	0.329	0.269	0.322	0.321	0.351	0.281	0.342	0.325	0.366
	720	0.351	0.381	0.348	0.378	0.384	0.410	0.357	0.376	0.366	0.382	0.362	0.385	0.363	0.382	0.408	0.403	0.397	0.421	0.421	0.415
	Avg.	0.256	0.317	0.248	0.311	0.267	0.327	0.254	0.311	0.261	0.316	0.255	0.315	0.251	0.309	0.291	0.333	0.267	0.332	0.305	0.349
Illness	24	1.409	0.754	-	-	1.613	0.834	1.869	0.823	1.792	0.807	1.319	0.754	2.255	1.017	2.317	0.934	2.215	1.081	3.228	1.260
	36	1.290	0.745	-	-	1.316	0.750	1.853	0.854	1.833	0.833	1.430	0.834	2.121	0.950	1.972	0.920	1.963	0.963	2.679	1.080
	48	1.499	0.810	-	-	1.548	0.818	1.886	0.855	2.269	1.012	1.553	0.815	2.187	1.007	2.238	0.940	2.130	1.024	2.622	1.078
	60	1.428	0.773	-	-	1.450	0.783	1.877	0.877	2.177	0.925	1.470	0.788	2.185	0.997	2.027	0.928	2.368	1.096	2.857	1.157
	Avg.	1.407	0.771	-	-	1.482	0.796	1.871	0.852	2.018	0.894	1.443	0.798	2.187	0.992	2.139	0.931	2.169	1.041	2.847	1.144
Electricity	96	0.126	0.213	0.142	0.245	0.127	0.217	0.141	0.239	0.137	0.233	0.129	0.222	0.128	0.223	0.168	0.272	0.140	0.237	0.193	0.308
	192	0.145	0.237	0.157	0.260	0.148	0.237	0.158	0.253	0.152	0.247	0.157	0.240	0.144	0.237	0.184	0.289	0.153	0.249	0.201	0.315
	336	0.162	0.254	0.174	0.276	0.163	0.253	0.172	0.266	0.169	0.267	0.163	0.259	0.160	0.254	0.198	0.300	0.169	0.267	0.214	0.329
	720	0.197	0.286	0.214	0.308	0.199	0.293	0.207	0.293	0.200	0.290	0.197	0.290	0.198	0.287	0.220	0.320	0.203	0.301	0.246	0.355
	Avg.	0.158	0.248	0.172	0.272	0.159	0.250	0.170	0.263	0.165	0.259	0.162	0.253	0.158	0.250	0.193	0.295	0.166	0.264	0.214	0.327
Weather	96	0.143	0.195	0.148	0.200	0.146	0.191	0.148	0.188	0.155	0.199	0.149	0.198	0.167	0.221	0.172	0.220	0.176	0.237	0.217	0.296
	192	0.187	0.242	0.193	0.240	0.194	0.238	0.192	0.230	0.223	0.261	0.194	0.241	0.212	0.258	0.219	0.261	0.220	0.282	0.276	0.336
	336	0.237	0.273	0.243	0.281	0.243	0.275	0.246	0.273	0.251	0.279	0.245	0.282	0.260	0.293	0.280	0.306	0.265	0.319	0.339	0.380
	720	0.302	0.315	0.312	0.332	0.318	0.328	0.320	0.328	0.345	0.342	0.314	0.334	0.328	0.339	0.365	0.359	0.333	0.362	0.403	0.428
	Avg.	0.217	0.256	0.224	0.263	0.225	0.258	0.227	0.255	0.244	0.270	0.226	0.264	0.242	0.278	0.259	0.287	0.249	0.300	0.309	0.360
Traffic	96	0.344	0.237	0.393	0.290	0.346	0.232	0.396	0.264	0.392	0.267	0.360	0.249	0.397	0.278	0.593	0.321	0.410	0.282	0.587	0.366
	192	0.363	0.249	0.405	0.296	0.376	0.245	0.412	0.268	0.409	0.271	0.379	0.256	0.411	0.283	0.617	0.336	0.423	0.287	0.604	0.373
	336	0.387	0.256	0.420	0.305	0.389	0.252	0.421	0.273	0.434	0.296	0.392	0.264	0.424	0.289	0.629	0.336	0.436	0.296	0.621	0.383
	720	0.433	0.284	0.459	0.323	0.432	0.293	0.455	0.291	0.451	0.291	0.432	0.286	0.450	0.305	0.640	0.350	0.466	0.315	0.626	0.382
	Avg.	0.382	0.257	0.419	0.304	0.386	0.256	0.421	0.274	0.422	0.281	0.391	0.264	0.421	0.289	0.620	0.336	0.434	0.295	0.610	0.376

IV Experiment

DMMV-S与DMMV-A对比

可以看出，对于同一个输入序列，DMMV-A分解出的周期性成分（e）非常平滑且周期性清晰，其趋势成分（f）则包含了主要的长期走向。相比之下，DMMV-S的分解结果则不够干净，周期性成分（b）中混杂了许多噪声，这无疑增加了视觉预测器的学习难度。这直观地解释了为何DMMV-A的自适应分解更为优越



IV Experiment

消融实验

为了验证模型中每个组件的必要性，设置了以下消融实验

(a) 和 (b)：将数值预测器换成更复杂的变换器或者将视觉预测器MAE换成SimMIM，

(c)：将门控融合换成简单的相加

(d) 和 (e)：分别用“无掩码”和“随机掩码”替代BCMASK

(g)：去掉整个回溯-残差分解机制，让两个预测器都处理原始输入

Dataset (→)	ETTh1		ETTm1		Illness		Weather	
Method (↓), Metric (→)	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
DMMV-A	0.395	0.414	0.340	0.371	1.407	0.771	0.217	0.256
(a) $f_{\text{num}}(\cdot) \rightarrow \text{Transformer}$ Improvement	0.407 -3.04%	0.421 -1.69%	0.339 +0.29%	0.372 -0.27%	1.442 -2.49%	0.786 -1.95%	0.219 -0.92%	0.260 -1.56%
(b) $f_{\text{vis}}(\cdot) \rightarrow \text{SimMIM}$ Improvement	0.407 -3.04%	0.415 -0.24%	0.345 -1.47%	0.377 -1.62%	1.649 -17.20%	0.814 -5.58%	0.227 -4.61%	0.261 -1.95%
(c) Gate \rightarrow Sum Improvement	0.414 -4.81%	0.427 -3.14%	0.352 -3.53%	0.383 -3.23%	1.606 -14.14%	0.863 -11.93%	0.233 -7.37%	0.278 -8.59%
(d) BCMASK \rightarrow No mask Improvement	0.426 -7.85%	0.441 -6.52%	0.349 -2.65%	0.377 -1.62%	1.493 -6.11%	0.828 -7.39%	0.221 -1.84%	0.267 -4.30%
(e) BCMASK \rightarrow Random mask Improvement	0.394 0.25%	0.414 0.00%	0.340 0.00%	0.372 -0.27%	1.472 -4.62%	0.829 -7.52%	0.223 -2.76%	0.262 -2.34%
(f) Freeze $f_{\text{vis}}(\cdot)$ Improvement	0.431 -9.11%	0.428 -3.38%	0.358 -5.29%	0.380 -2.43%	1.442 -2.49%	0.773 -0.26%	0.246 -13.36%	0.288 -12.50%
(g) W/o decomposition Improvement	0.408 -3.29%	0.424 -2.42%	0.338 0.59%	0.373 -0.54%	1.712 -21.68%	0.903 -17.12%	0.219 -0.92%	0.268 -4.69%

V Conclusion



V Conclusion

本文从理论和实验上证明，通过一个精心设计的分解框架，可以有效结合数值和视觉两种模态的优势，显著提升长时序预测的性能。其提出的DMMV-A模型通过新颖的自适应分解机制，成功克服了现有LVM方法中的周期性偏置问题，在多个基准测试中刷新了记录。这项研究不仅为时序预测任务提供了一个强大的新工具，也为未来如何将不同模态的大模型应用于时序分析领域，开辟了富有前景的研究方向。

谢谢！

