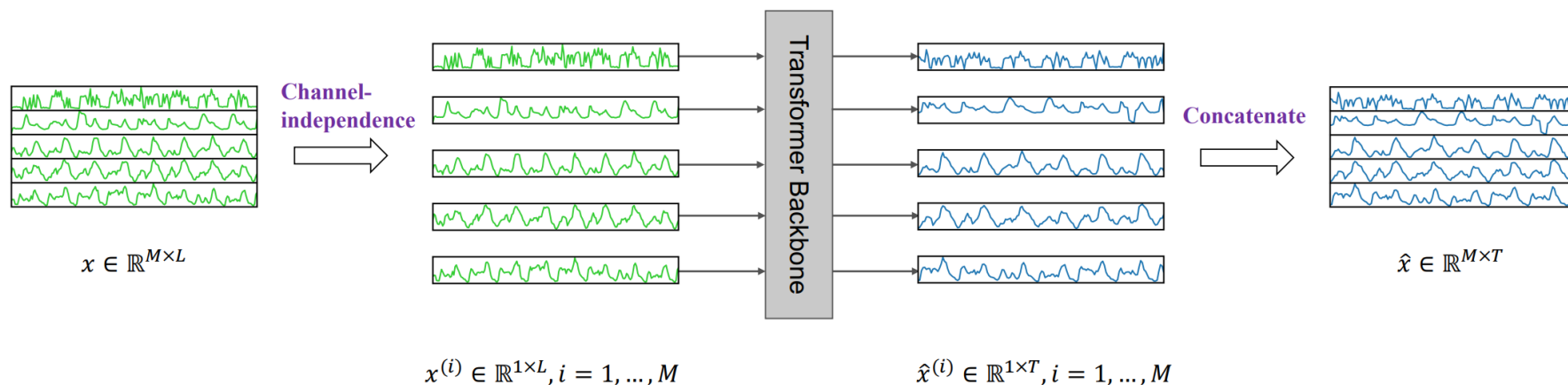
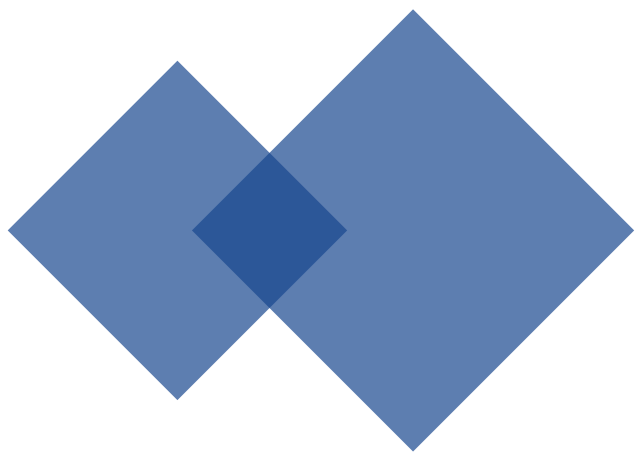


PatchTST

A TIME SERIES IS WORTH 64
WORDS: LONG-TERM
FORECASTING WITH
TRANSFORMERS

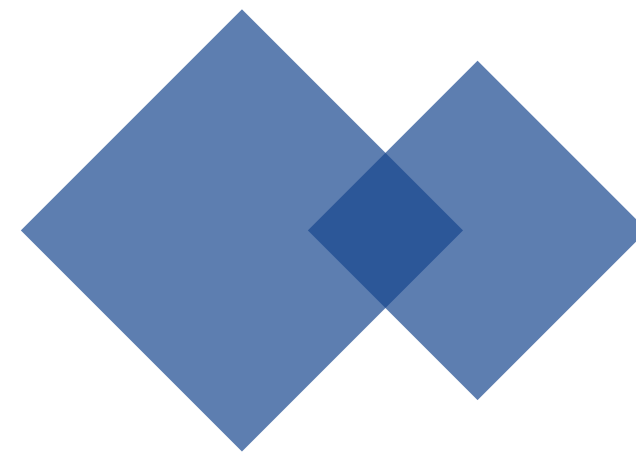


(a) PatchTST Model Overview



PatchTST

A TIME SERIES IS WORTH 64
WORDS: LONG-TERM
FORECASTING WITH
TRANSFORMERS



24.3.19

Presented by Yyyq



➤ **Transformer-based model for**

- 多变量时序预测
- 自监督表示学习

➤ **channel-independent patch time series Transformer (PatchTST)**

- Input Token: 将时间序列分割为子序列级patch
- 通道独立: 每个通道包含单个单变量时间序列

Input Token（原）：将每一个时间点视为一个token

Patch: 时间序列被分成若干个时间段（segment），每个时间段视为一个token



关于通道混合 Channel-mixing 和 通道独立 Channel-independence

通道混合强调不同通道之间的相关性和交互性建模，提高模型的表达能力和泛化能力。

- **Transformer的通道混合**

- 1) 自注意力机制：每个位置的表示都是所有位置的加权组合
- 2) 多头注意力机制：不同的注意力头中学习不同的特征表示，进行合并

- **CNN 的通道独立**

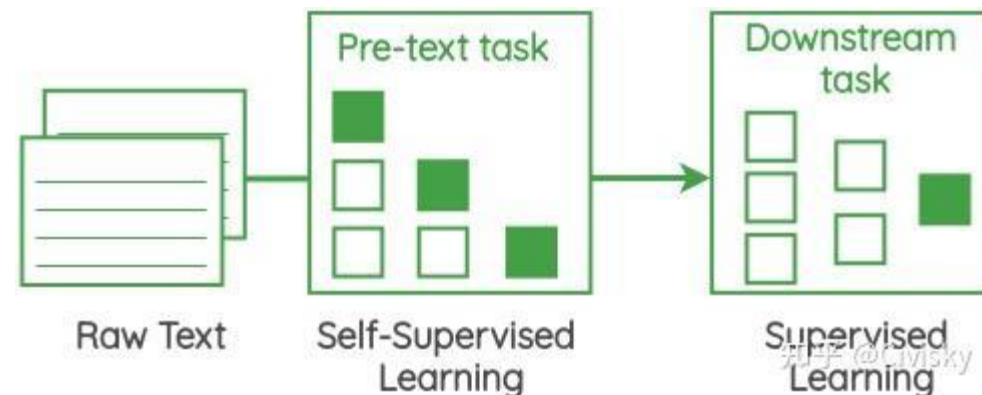
卷积操作是局部连接的，每个通道的特征表示仅由局部区域的输入数据决定，与其他通道的特征表示是独立的。

- **Linear的通道独立**

全连接结构，每个神经元对输入数据的权重是独立学习的，不考虑输入数据的通道关系。



- **Transformer-based model for**
 - 多变量时序预测
 - 自监督表示学习
- **channel-independent patch time series Transformer (PatchTST)**
 - Input Token: 将时间序列分割为子序列级patch
 - 通道独立: 每个通道包含单个单变量时间序列
- **模型优势**
 - 时间和空间复杂度的降低: $N = L \rightarrow N \approx L / S$
 - 可以关注更长的历史序列: 精准长时预测
 - 表示学习的能力: 捕获数据的抽象表示



“表示学习的能力：捕获数据的抽象表示”

通过自监督学习（self-supervised learning）学习有用的表示，进一步服务于下游任务

自监督学习的 **Pretrain - Finetune** 流程：

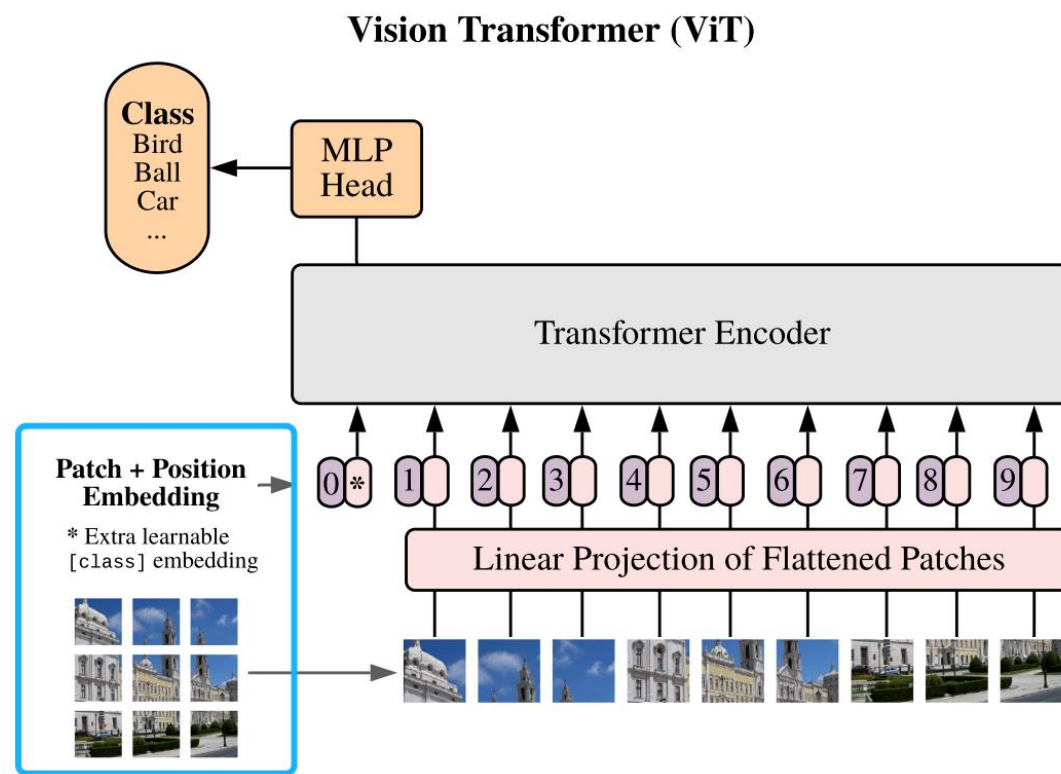
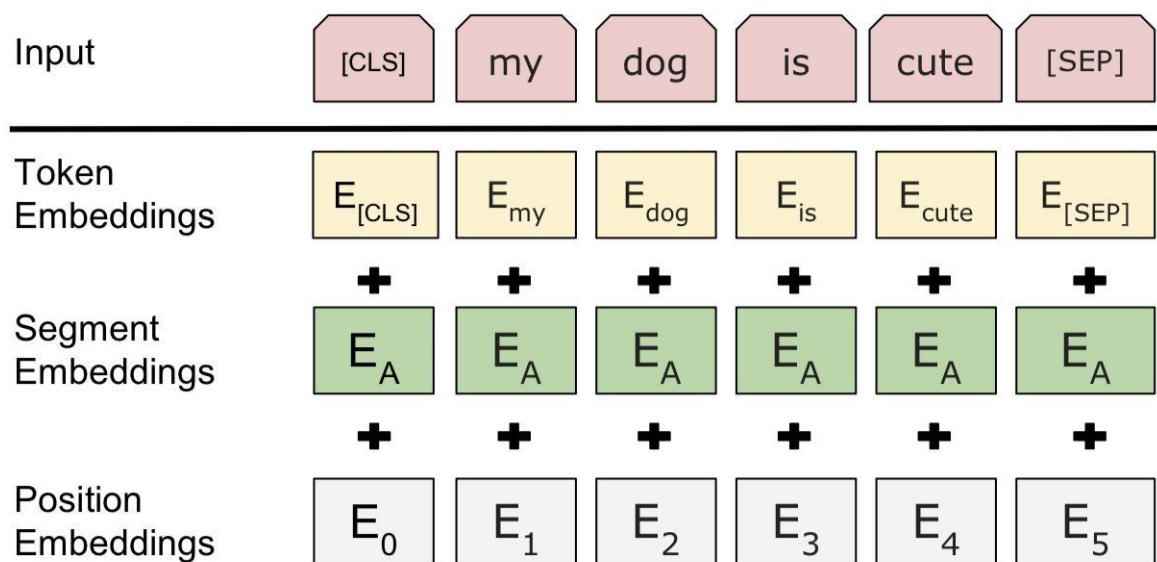
- 从大量的无标签数据中通过“辅助任务pretext”来训练网络 → 得到预训练的模型
- 对于新的下游任务，在学习到的参数的基础上 → 微调模型

Transformer 作为一个基础模型，其在表示学习上的潜力没有被完全开发

Models	L	N	patch	method	MSE
Channel-independent PatchTST	96	96		down-sampled	0.518
	380	96			0.447
	336	336			0.397
	336	42	✓	self-supervised	0.367
	336	42	✓		0.349
Channel-mixing FEDFormer DLinear	336	336			0.597
	336	336			0.410

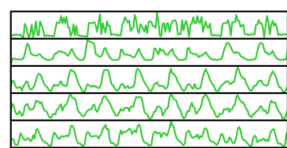
➤ Patch in Transformer-based Models

- NLP领域——BERT：基于子词的标记化，而非字符
- CV领域——ViT (Vision Transformer)：将输入图像分割成均匀大小的图像块



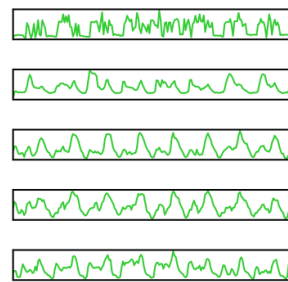
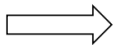
03

算法实现



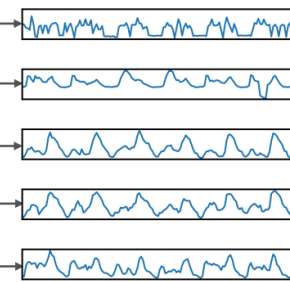
$$x \in \mathbb{R}^{M \times L}$$

Channel-independence



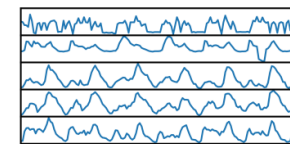
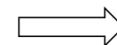
$$x^{(i)} \in \mathbb{R}^{1 \times L}, i = 1, \dots, M$$

Transformer Backbone



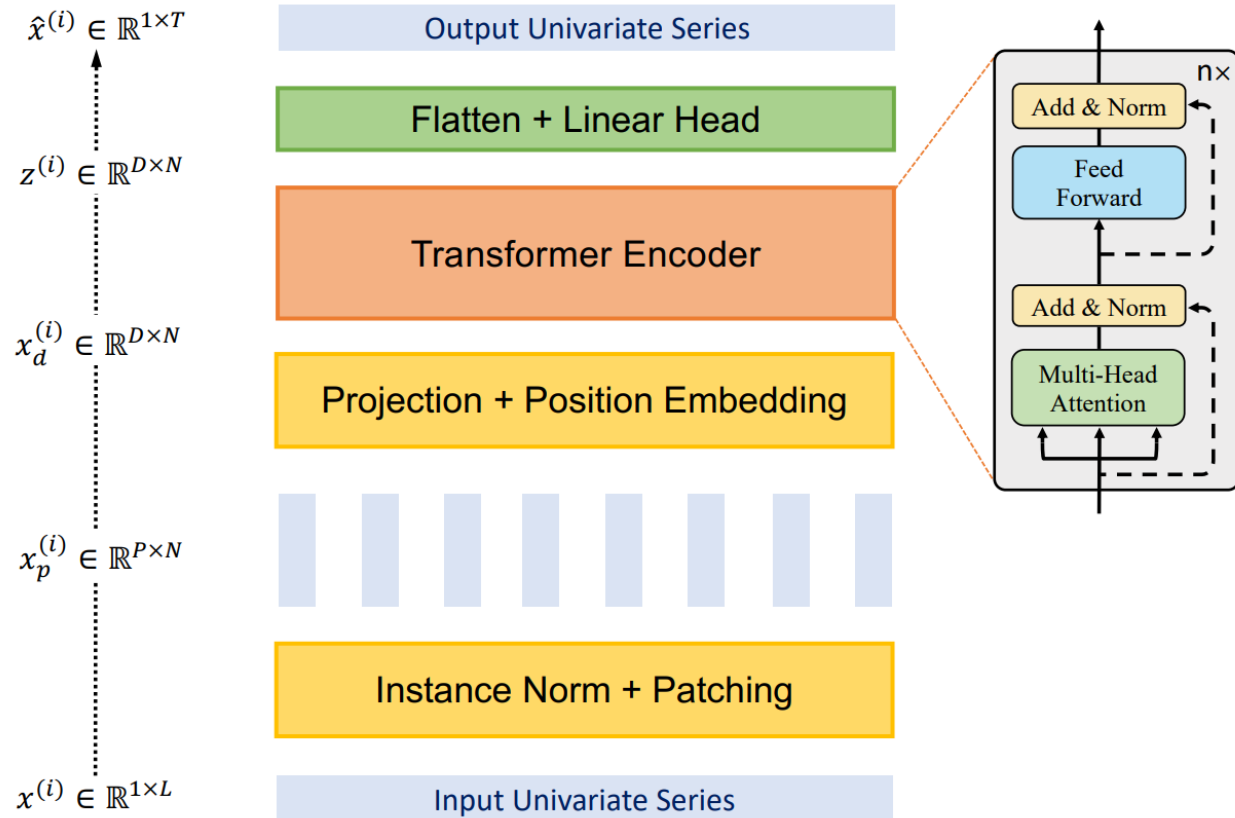
$$\hat{x}^{(i)} \in \mathbb{R}^{1 \times T}, i = 1, \dots, M$$

Concatenate



$$\hat{x} \in \mathbb{R}^{M \times T}$$

(a) PatchTST Model Overview





- 用过去L步预测未来的T步, 输入是M个变量

$$(\mathbf{x}_1, \dots, \mathbf{x}_L) \quad \rightarrow \quad (\mathbf{x}_{L+1}, \dots, \mathbf{x}_{L+T})$$

- 通道独立 → 每个序列单独看, transformer主干将提供每个序列的预测结果

$$\mathbf{x}^{(i)} \in \mathbb{R}^{1 \times L}, \quad \mathbf{x}_{1:L}^{(i)} = (x_1^{(i)}, \dots, x_L^{(i)}) \text{ where } i = 1, \dots, M.$$

$$\hat{\mathbf{x}}^{(i)} = (\hat{x}_{L+1}^{(i)}, \dots, \hat{x}_{L+T}^{(i)}) \in \mathbb{R}^{1 \times T}$$



- Patch长度 (P)
- Stride步长 (S) : patch可以重叠或非重叠
- Token个数 (N) : $N = \lfloor \frac{(L-P)}{S} \rfloor + 2.$

$$\mathbf{x}_p^{(i)} \in \mathbb{R}^{P \times N}$$

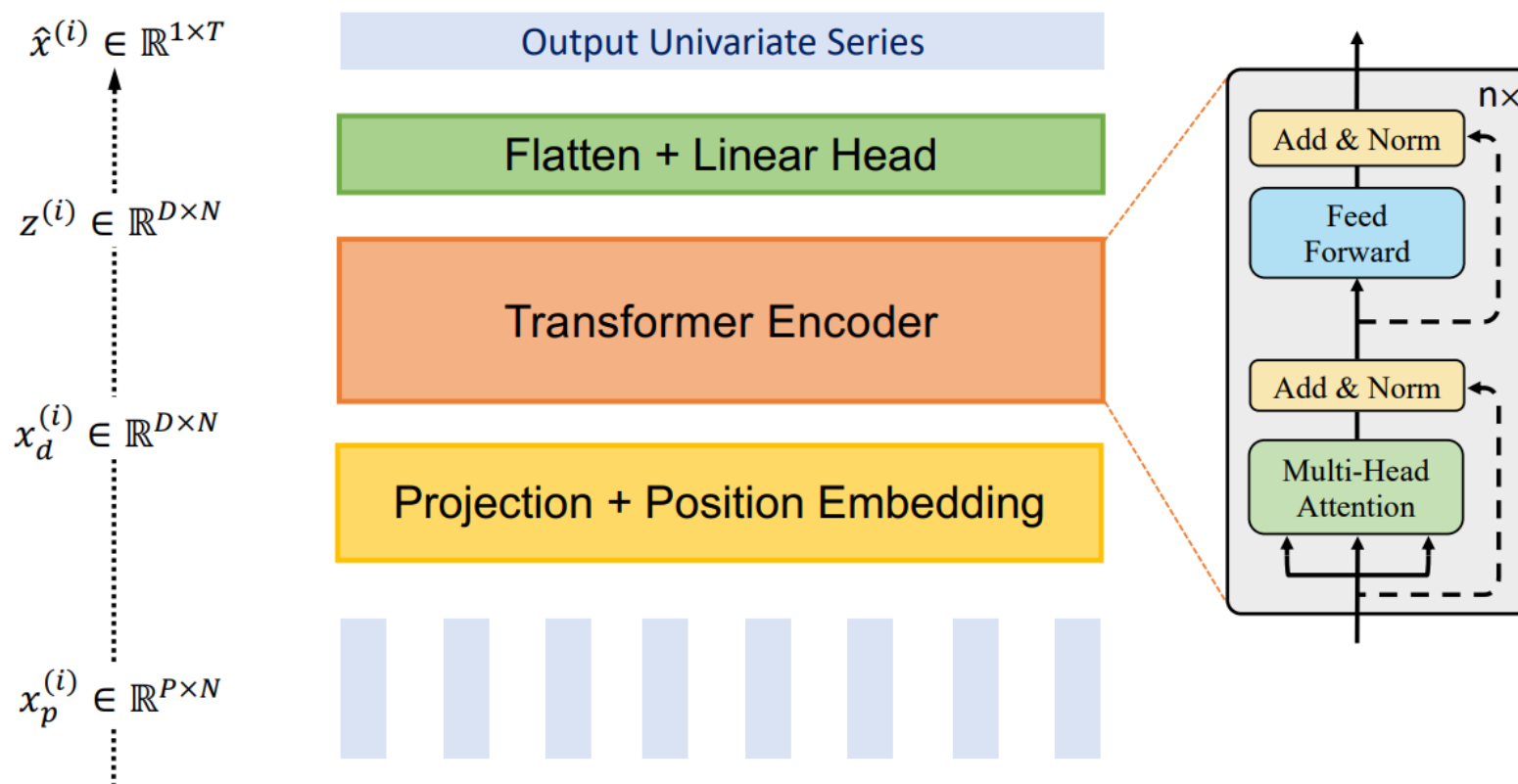
- 输入的token数量减少: $L \rightarrow L/S$
- 而注意力机制的内存和计算复杂度: 成平方减少

03



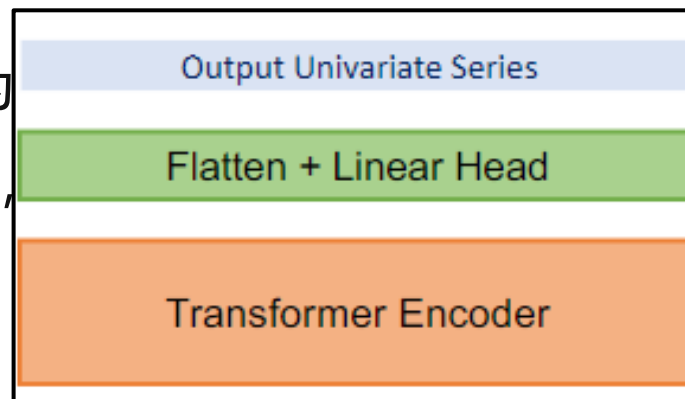
算法实现: Transformer Encoder

- 获得每个单变量的预测结果



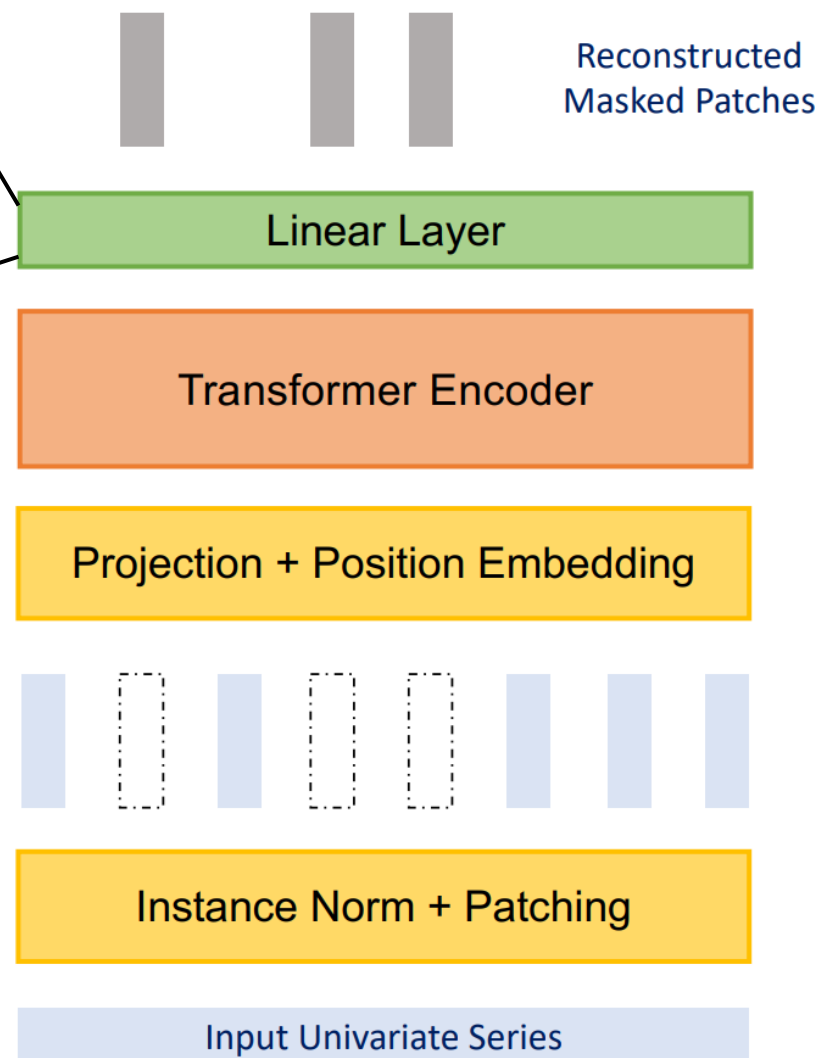
➤ 预训练辅助任务

- 将每个输入序列划分为
- 随机掩码，掩蔽patch
- 重建被遮蔽的patch



➤ 解决了两个问题

- 掩蔽数据点，插值导致无法学习到重要的抽象表示
- $(L \cdot D) \times (M \cdot T) \rightarrow (D \times P)$





Datasets	Weather	Traffic	Electricity	ILI	ETTh1	ETTh2	ETTm1	ETTm2
Features	21	862	321	7	7	7	7	7
Timesteps	52696	17544	26304	966	17420	17420	69680	69680

大型数据集：Weather, Traffic, Electricity

回望窗口:

FEDformer、Autoformer、Informer:

$L \in \{24, 48, 96, 192, 336, 720\}$ 选最佳

Dlinear: $L = 336$

PatchTST: $P=16, S=8$

PatchTST/64: $L = 512, N = 64$

PatchTST/42: $L = 336, N = 42$

(PatchTST/42与Dlinear回望窗口相同)

Models		PatchTST/64		PatchTST/42		DLinear		FEDformer		Autoformer		Informer		Pyraformer		LogTrans	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather	96	0.149	0.198	0.152	0.199	0.176	0.237	0.238	0.314	0.249	0.329	0.354	0.405	0.896	0.556	0.458	0.490
	192	0.194	0.241	0.197	0.243	0.220	0.282	0.275	0.329	0.325	0.370	0.419	0.434	0.622	0.624	0.658	0.589
	336	0.245	0.282	0.249	0.283	0.265	0.319	0.339	0.377	0.351	0.391	0.583	0.543	0.739	0.753	0.797	0.652
	720	0.314	0.334	0.320	0.335	0.323	0.362	0.389	0.409	0.415	0.426	0.916	0.705	1.004	0.934	0.869	0.675
Traffic	96	0.360	0.249	0.367	0.251	0.410	0.282	0.576	0.359	0.597	0.371	0.733	0.410	2.085	0.468	0.684	0.384
	192	0.379	0.256	0.385	0.259	0.423	0.287	0.610	0.380	0.607	0.382	0.777	0.435	0.867	0.467	0.685	0.390
	336	0.392	0.264	0.398	0.265	0.436	0.296	0.608	0.375	0.623	0.387	0.776	0.434	0.869	0.469	0.734	0.408
	720	0.432	0.286	0.434	0.287	0.466	0.315	0.621	0.375	0.639	0.395	0.827	0.466	0.881	0.473	0.717	0.396
Electricity	96	0.129	0.222	0.130	0.222	0.140	0.237	0.186	0.302	0.196	0.313	0.304	0.393	0.386	0.449	0.258	0.357
	192	0.147	0.240	0.148	0.240	0.153	0.249	0.197	0.311	0.211	0.324	0.327	0.417	0.386	0.443	0.266	0.368
	336	0.163	0.259	0.167	0.261	0.169	0.267	0.213	0.328	0.214	0.327	0.333	0.422	0.378	0.443	0.280	0.380
	720	0.197	0.290	0.202	0.291	0.203	0.301	0.233	0.344	0.236	0.342	0.351	0.427	0.376	0.445	0.283	0.376
ILI	24	1.319	0.754	1.522	0.814	2.215	1.081	2.624	1.095	2.906	1.182	4.657	1.449	1.420	2.012	4.480	1.444
	36	1.579	0.870	1.430	0.834	1.963	0.963	2.516	1.021	2.585	1.038	4.650	1.463	7.394	2.031	4.799	1.467
	48	1.553	0.815	1.673	0.854	2.130	1.024	2.505	1.041	3.024	1.145	5.004	1.542	7.551	2.057	4.800	1.468
	60	1.470	0.788	1.529	0.862	2.368	1.096	2.742	1.122	2.761	1.114	5.071	1.543	7.662	2.100	5.278	1.560
ETTh1	96	0.370	0.400	0.375	0.399	0.375	0.399	0.376	0.415	0.435	0.446	0.941	0.769	0.664	0.612	0.878	0.740
	192	0.413	0.429	0.414	0.421	0.405	0.416	0.423	0.446	0.456	0.457	1.007	0.786	0.790	0.681	1.037	0.824
	336	0.422	0.440	0.431	0.436	0.439	0.443	0.444	0.462	0.486	0.487	1.038	0.784	0.891	0.738	1.238	0.932
	720	0.447	0.468	0.449	0.466	0.472	0.490	0.469	0.492	0.515	0.517	1.144	0.857	0.963	0.782	1.135	0.852
ETTh2	96	0.274	0.337	0.274	0.336	0.289	0.353	0.332	0.374	0.332	0.368	1.549	0.952	0.645	0.597	2.116	1.197
	192	0.341	0.382	0.339	0.379	0.383	0.418	0.407	0.446	0.426	0.434	3.792	1.542	0.788	0.683	4.315	1.635
	336	0.329	0.384	0.331	0.380	0.448	0.465	0.400	0.447	0.477	0.479	4.215	1.642	0.907	0.747	1.124	1.604
	720	0.379	0.422	0.379	0.422	0.605	0.551	0.412	0.469	0.453	0.490	3.656	1.619	0.963	0.783	3.188	1.540
ETTm1	96	0.293	0.346	0.290	0.342	0.299	0.343	0.326	0.390	0.510	0.492	0.626	0.560	0.543	0.510	0.600	0.546
	192	0.333	0.370	0.332	0.369	0.335	0.365	0.365	0.415	0.514	0.495	0.725	0.619	0.557	0.537	0.837	0.700
	336	0.369	0.392	0.366	0.392	0.369	0.386	0.392	0.425	0.510	0.492	1.005	0.741	0.754	0.655	1.124	0.832
	720	0.416	0.420	0.420	0.424	0.425	0.421	0.446	0.458	0.527	0.493	1.133	0.845	0.908	0.724	1.153	0.820
ETTm2	96	0.166	0.256	0.165	0.255	0.167	0.260	0.180	0.271	0.205	0.293	0.355	0.462	0.435	0.507	0.768	0.642
	192	0.223	0.296	0.220	0.292	0.224	0.303	0.252	0.318	0.278	0.336	0.595	0.586	0.730	0.673	0.989	0.757
	336	0.274	0.329	0.278	0.329	0.281	0.342	0.324	0.364	0.343	0.379	1.270	0.871	1.201	0.845	1.334	0.872
	720	0.362	0.385	0.367	0.385	0.397	0.421	0.410	0.420	0.414	0.419	3.001	1.267	3.625	1.451	3.048	1.328

Models		Fine-tuning		PatchTST		Sup.		DLinear		FEDformer		Autoformer		Informer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather	96	0.144	0.193	0.158	0.209	<u>0.152</u>	<u>0.199</u>	0.176	0.237	0.238	0.314	0.249	0.329	0.354	0.405
	192	0.190	0.236	0.203	0.249	<u>0.197</u>	<u>0.243</u>	0.220	0.282	0.275	0.329	0.325	0.370	0.419	0.434
	336	0.244	0.280	0.251	0.285	<u>0.249</u>	<u>0.283</u>	0.265	0.319	0.339	0.377	0.351	0.391	0.583	0.543
	720	0.320	0.335	0.321	0.336	0.320	0.335	0.323	0.362	0.389	0.409	0.415	0.426	0.916	0.705
Traffic	96	0.352	0.244	0.399	0.294	<u>0.367</u>	<u>0.251</u>	0.410	0.282	0.576	0.359	0.597	0.371	0.733	0.410
	192	0.371	0.253	0.412	0.298	<u>0.385</u>	<u>0.259</u>	0.423	0.287	0.610	0.380	0.607	0.382	0.777	0.435
	336	0.381	0.257	0.425	0.306	<u>0.398</u>	<u>0.265</u>	0.436	0.296	0.608	0.375	0.623	0.387	0.776	0.434
	720	0.425	0.282	0.460	0.323	<u>0.434</u>	<u>0.287</u>	0.466	0.315	0.621	0.375	0.639	0.395	0.827	0.466
Electricity	96	0.126	0.221	0.138	0.237	<u>0.130</u>	0.222	0.140	0.237	0.186	0.302	0.196	0.313	0.304	0.393
	192	0.145	0.238	0.156	0.252	<u>0.148</u>	<u>0.240</u>	0.153	0.249	0.197	0.311	0.211	0.324	0.327	0.417
	336	0.164	0.256	0.170	0.265	<u>0.167</u>	<u>0.261</u>	0.169	0.267	0.213	0.328	0.214	0.327	0.333	0.422
	720	0.193	0.291	0.208	0.297	<u>0.202</u>	0.291	0.203	0.301	0.233	0.344	0.236	0.342	0.351	0.427

Patch不重叠, $L = 512$, $P = 12$, 掩码率40%

- 自监督预训练：100个epochs
- 监督训练：
 - ① 线性探测：只训练模型头 20 个 epoch，同时冻结网络的其余部分
 - ② 微调：先线性探测10个epoch，再有监督地微调整个网络20个epoch

Models		PatchTST						DLinear		FEDformer		Autoformer		Informer	
		Fine-tuning		Lin. Prob.		Sup.									
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather	96	0.145	0.195	0.163	0.216	<u>0.152</u>	<u>0.199</u>	0.176	0.237	0.238	0.314	0.249	0.329	0.354	0.405
	192	0.193	0.243	0.205	0.252	<u>0.197</u>	0.243	0.220	0.282	0.275	0.329	0.325	0.370	0.419	0.434
	336	0.244	0.280	0.253	0.289	<u>0.249</u>	<u>0.283</u>	0.265	0.319	0.339	0.377	0.351	0.391	0.583	0.543
	720	0.321	0.337	0.320	<u>0.336</u>	0.320	0.335	0.323	0.362	0.389	0.409	0.415	0.426	0.916	0.705
Traffic	96	<u>0.388</u>	<u>0.273</u>	0.400	0.288	0.367	0.251	0.410	0.282	0.576	0.359	0.597	0.371	0.733	0.410
	192	<u>0.400</u>	<u>0.277</u>	0.412	0.293	0.385	0.259	0.423	0.287	0.610	0.380	0.607	0.382	0.777	0.435
	336	<u>0.408</u>	<u>0.280</u>	0.425	0.307	0.398	0.265	0.436	0.296	0.608	0.375	0.623	0.387	0.776	0.434
	720	<u>0.447</u>	<u>0.310</u>	0.457	0.317	0.434	0.287	0.466	0.315	0.621	0.375	0.639	0.395	0.827	0.466

Models		Fine-tuning	
		MSE	MAE
Weather	96	0.144	0.193
	192	0.190	0.236
	336	0.244	0.280
	720	0.320	0.335
Traffic	96	0.352	0.244
	192	0.371	0.253
	336	0.381	0.257
	720	0.425	0.282

电力数据集上预训练模型并在其他数据集上进行微调

05



实验2：表示学习——与SOTA的对比学习进行比较

Models		IMP.	PatchTST				BTSF		TS2Vec		TNC		TS-TCC	
			Transferred		Self-supervised									
Metrics		MSE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	24	42.3%	0.312	0.362	<u>0.322</u>	<u>0.369</u>	0.541	0.519	0.599	0.534	0.632	0.596	0.653	0.610
	48	44.7%	0.339	0.378	<u>0.354</u>	<u>0.385</u>	0.613	0.524	0.629	0.555	0.705	0.688	0.720	0.693
	168	34.5%	<u>0.424</u>	<u>0.437</u>	0.419	0.424	0.640	0.532	0.755	0.636	1.097	0.993	1.129	1.044
	336	48.5%	<u>0.472</u>	<u>0.472</u>	0.445	0.446	0.864	0.689	0.907	0.717	1.454	0.919	1.492	1.076
	720	48.8%	<u>0.508</u>	<u>0.507</u>	0.487	0.478	0.993	0.712	1.048	0.790	1.604	1.118	1.603	1.206

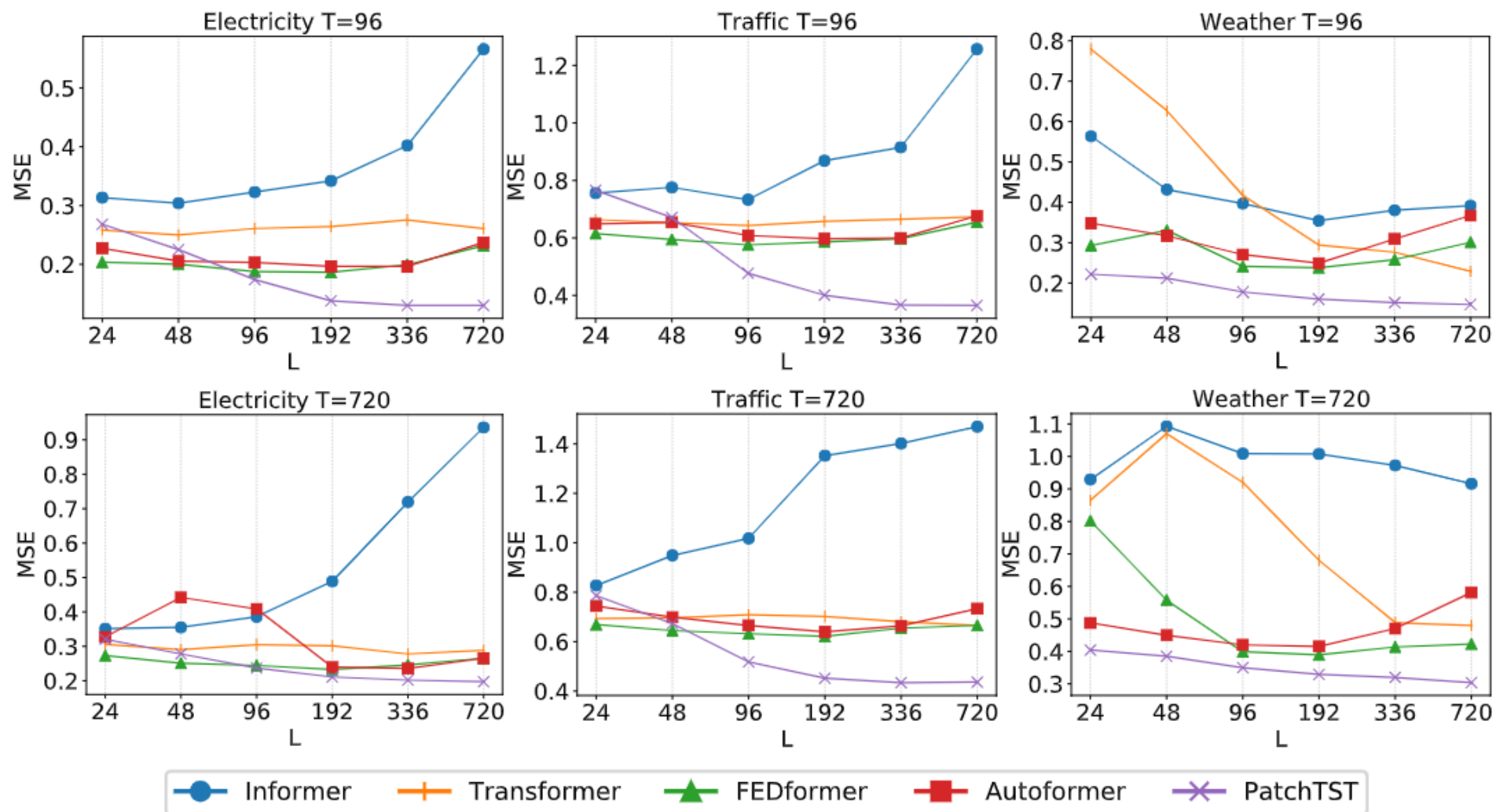
在traffic上训练，
在ETTh1微调

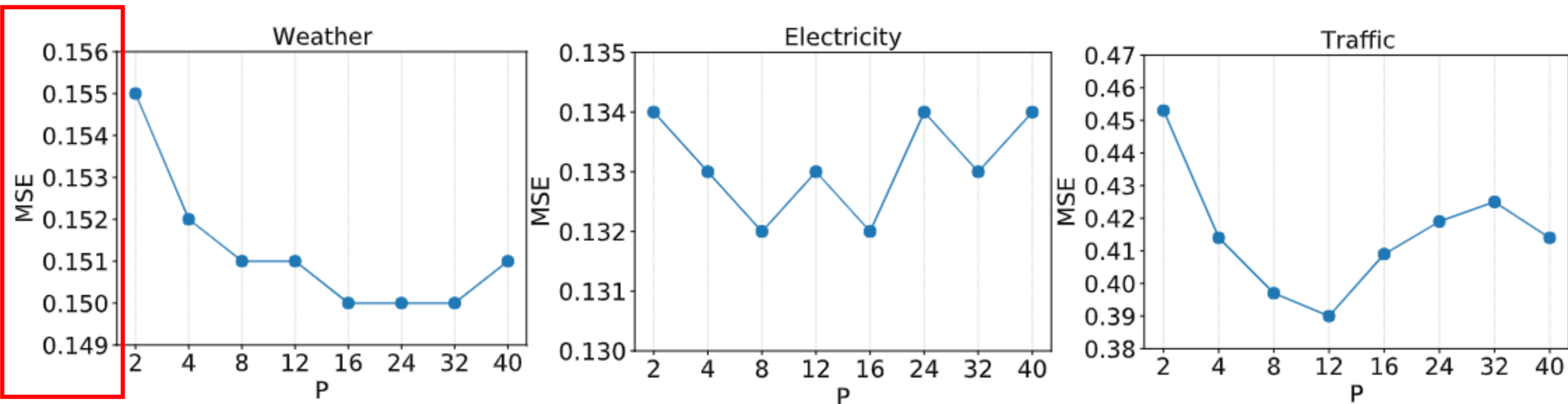
在ETTh1上预训练

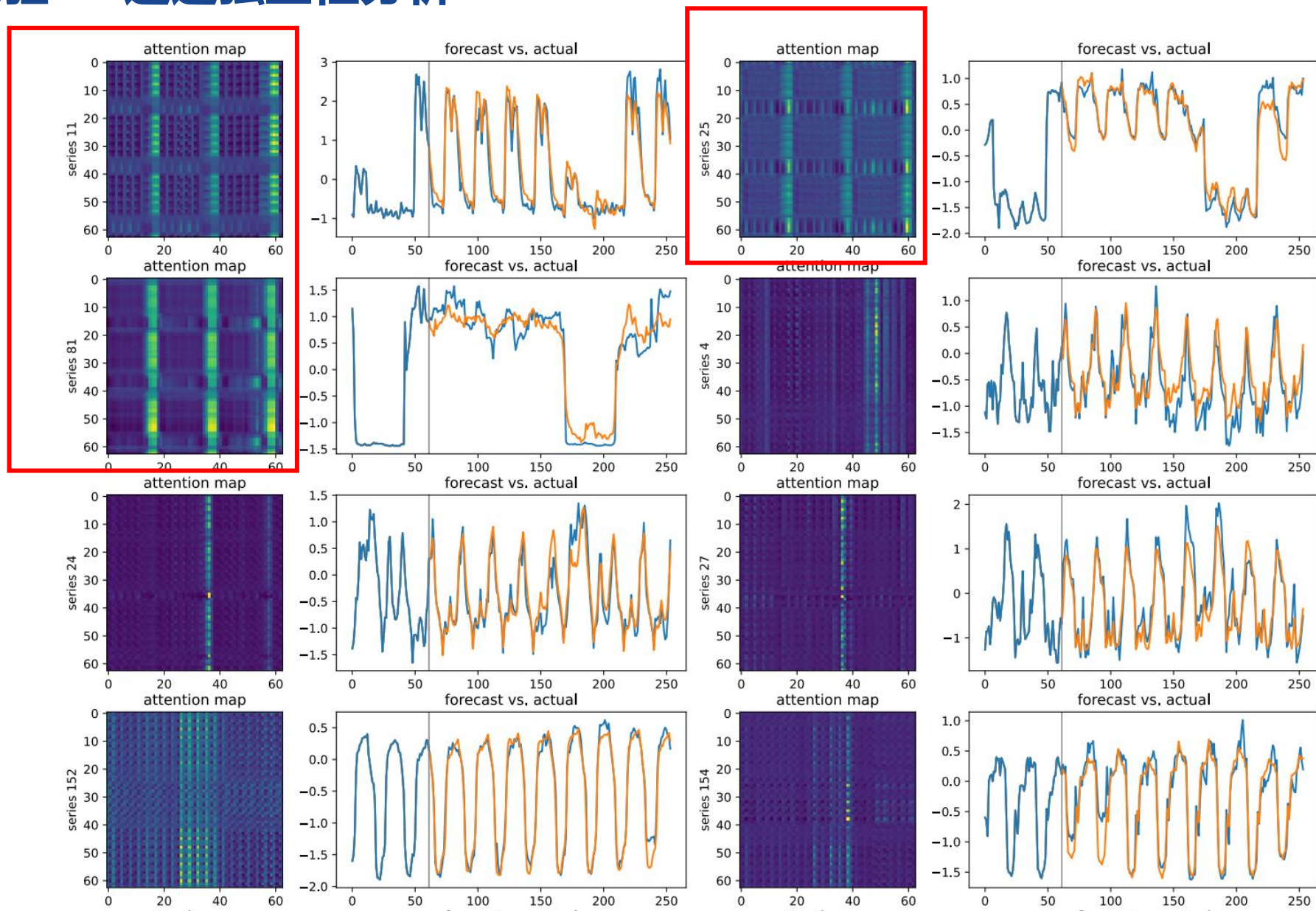


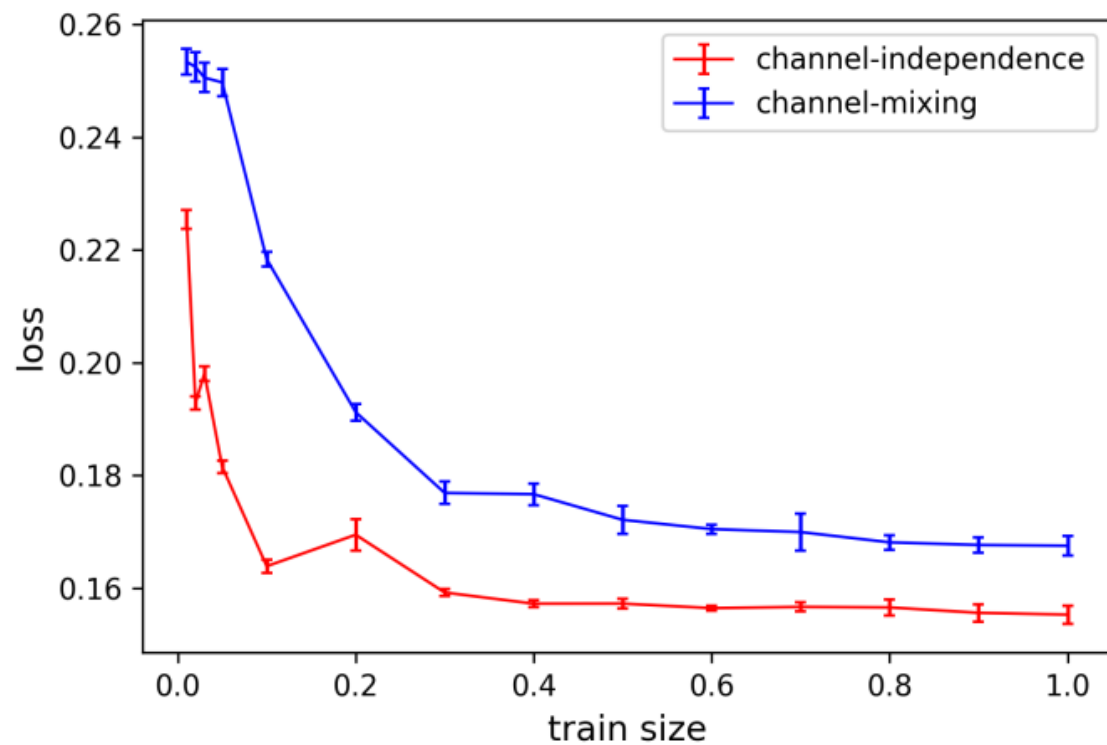
CI：只做通道独立，P：只做patch

Models		PatchTST								FEDformer	
		P+CI		CI		P		Original			
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather	96	0.152	0.199	0.164	0.213	0.168	0.223	0.177	0.236	0.238	0.314
	192	0.197	0.243	0.205	0.250	0.213	0.262	0.221	0.270	0.275	0.329
	336	0.249	0.283	0.255	0.289	0.266	0.300	0.271	0.306	0.339	0.377
	720	0.320	0.335	0.327	0.343	0.351	0.359	0.340	0.353	0.389	0.409
Traffic	96	0.367	0.251	0.397	0.271	0.595	0.376	-	-	0.576	0.359
	192	0.385	0.259	0.411	0.276	0.612	0.387	-	-	0.610	0.380
	336	0.398	0.265	0.423	0.282	0.651	0.391	-	-	0.608	0.375
	720	0.434	0.287	0.457	0.309	-	-	-	-	0.621	0.375
Electricity	96	0.130	0.222	0.136	0.231	0.196	0.307	0.205	0.318	0.186	0.302
	192	0.148	0.240	0.164	0.263	0.215	0.323	-	-	0.197	0.311
	336	0.167	0.261	0.168	0.262	0.228	0.338	-	-	0.213	0.328
	720	0.202	0.291	0.219	0.312	0.244	0.345	-	-	0.233	0.344

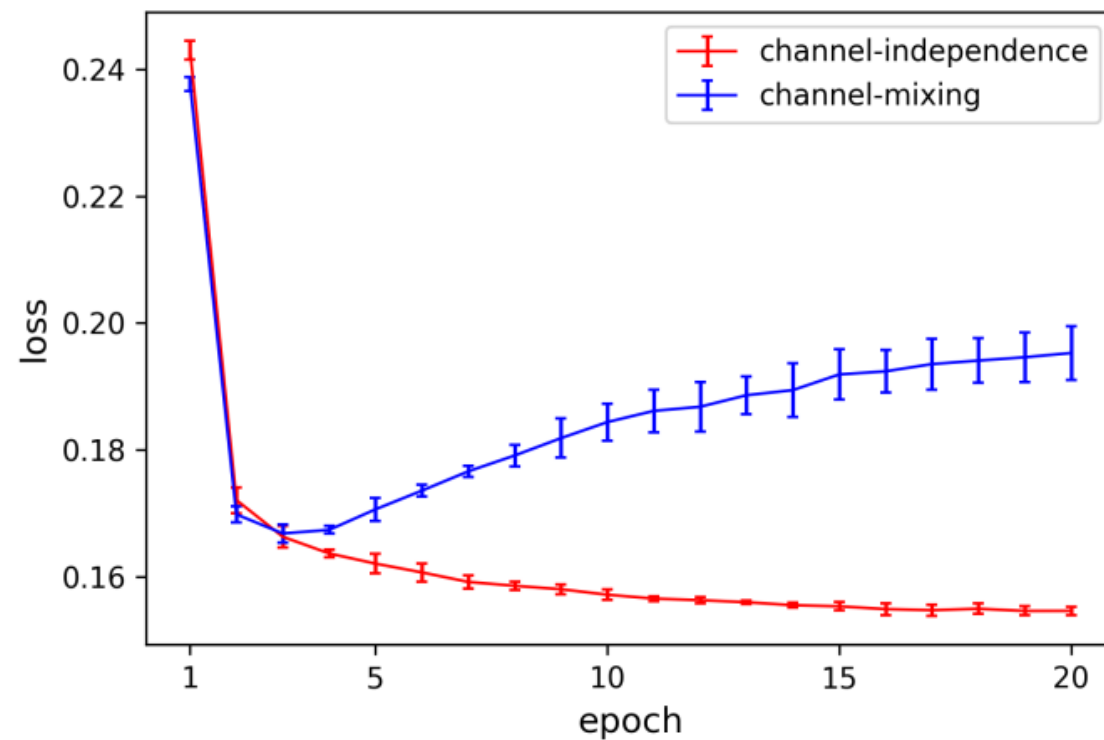








收敛得更快



不容易过拟合



谢谢观看

MANY THANKS !

23.3.19

