Towards CLEAN: Cycle Learning Energy-Emission Assessment Network

NeurIPS 2024 Workshop: Tackling Climate Change with Machine Learning

Yanming Guo, Qian Xiao, Jin Ma

yguo0337@uni.sydney.edu.au

https://github.com/Yvnminc/CLEAN

THE UNIVERSITY OF SYDNEY

Maynooth University National University of Ireland Maynooth

Limitation of Current Emission Dataset

Carbon Emission is one of most well discussed Greenhouse Gas in both academia and industry. It servers as an essential ecological measure to analysis the environmental impact of economic activities. Accurate emission data is a necessary condition to make climate policy. Despite various source of emission data reported by institutes, such as International Energy Agency (IEA). Several gaps and challenges in emission data persist:

	Spatial	Temporal
Coverage	Lack of Data in Rest of Worlds	Limited Time Span & Time Lags
Granularity	Limited on National Scales	Annually Basis Emission Data

These challenges can be categorized along spatial and temporal dimensions. Spatially, most datasets focus on major regions like China, the US, and the EU, while emission data for many other parts of the world, such as Singapore or Africa, are either incomplete or unavailable. Even within covered regions, there is limited geographical granularity. For instance, city-level emissions, such as those for New York or Shanghai, or state-level data, such as for California or Jiangsu Province, are hard to find.

Experiments and Results

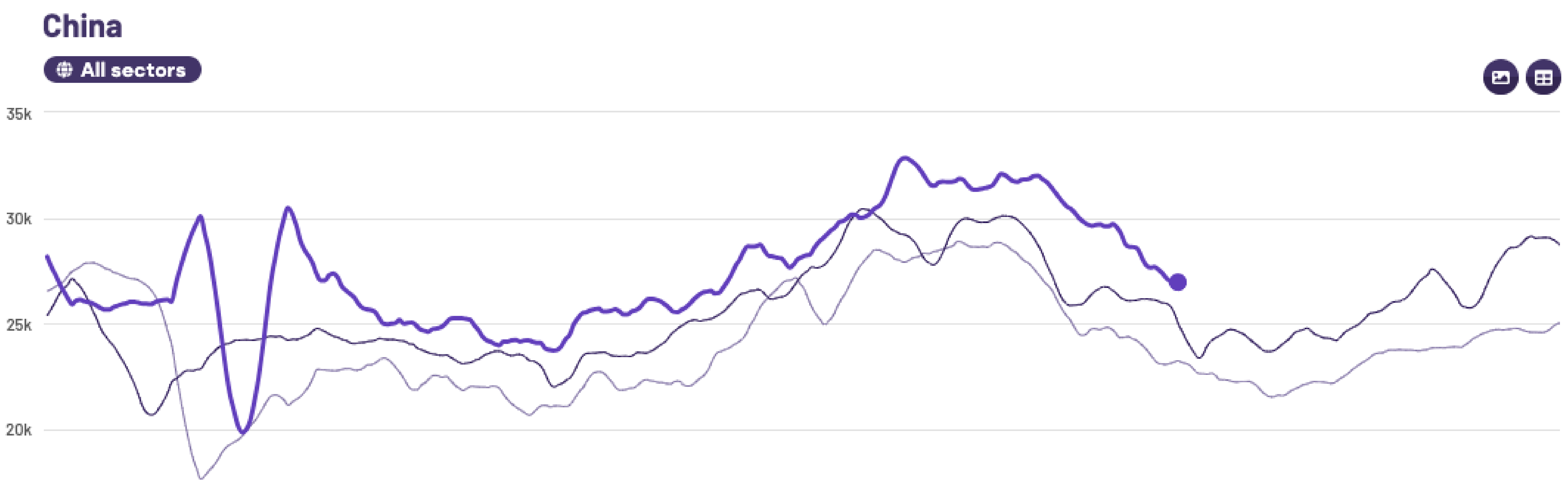
Table 2: Mean Absolute Percentage Error (MAPE) for Energy to Emission and Emission to Energy models with densification levels n. Here, n represents the number of synthesized countries; the more synthesized countries, the denser the training data space, and vice versa. The top results are in **bold**. It can be observed that deep learning models generally outperform traditional learning models, and those trained with the CLEAN framework and Cycle loss achieve the best performance.

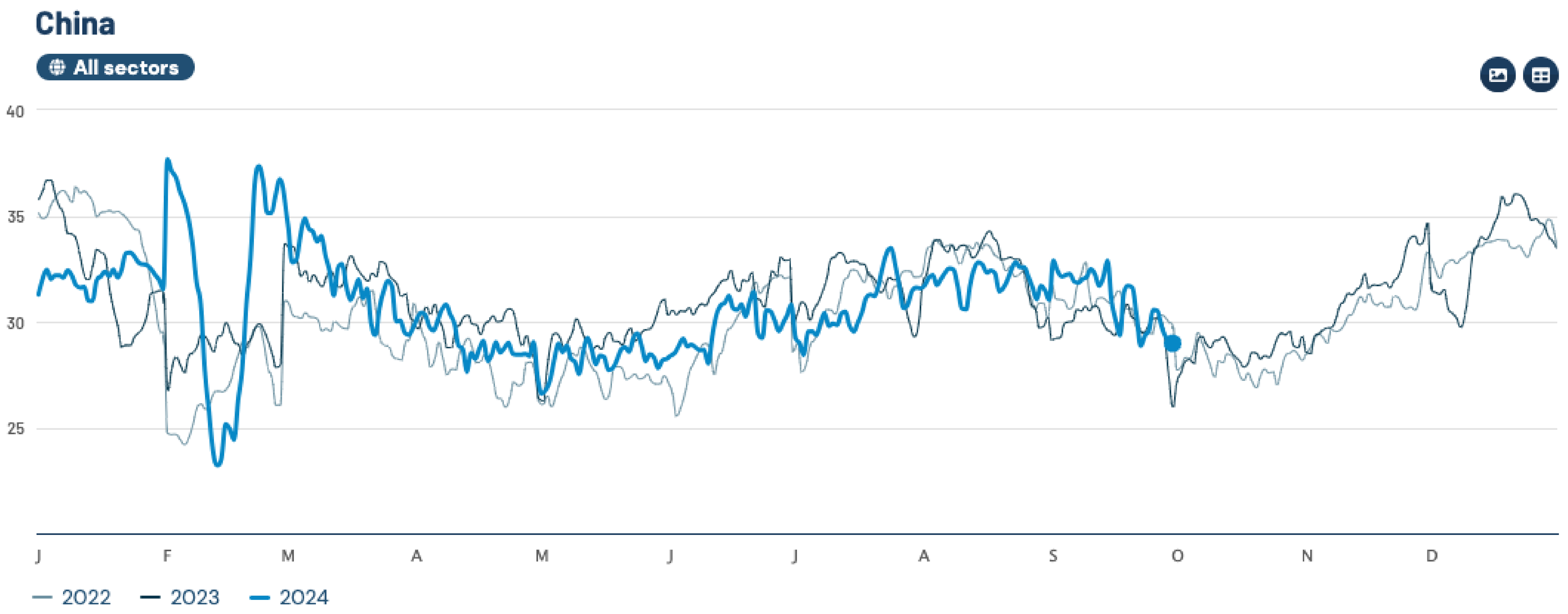
	CMEE_0		CMEE_{10}		CMEE_{100}	
Model	Emission	Energy	Emission	Energy	Emission	Energy
LR [8]	0.423	0.347	0.243	0.128	0.104	0.101
KNN [4]	0.309	0.461	0.349	0.316	0.070	0.102
DT [16]	0.768	1.256	0.402	0.450	0.094	0.128
XGB [2]	0.359	1.166	0.374	0.370	0.070	0.112
MLP [11]	0.136	0.157	0.183	0.112	0.056	**0.049**
CNN [11]	0.206	0.273	0.259	0.201	0.062	0.050
CLEAN_{MLP}	**0.118**	**0.111**	**0.125**	**0.110**	0.051	0.057
CLEAN_{CNN}	0.133	0.196	0.178	0.191	**0.048**	0.054

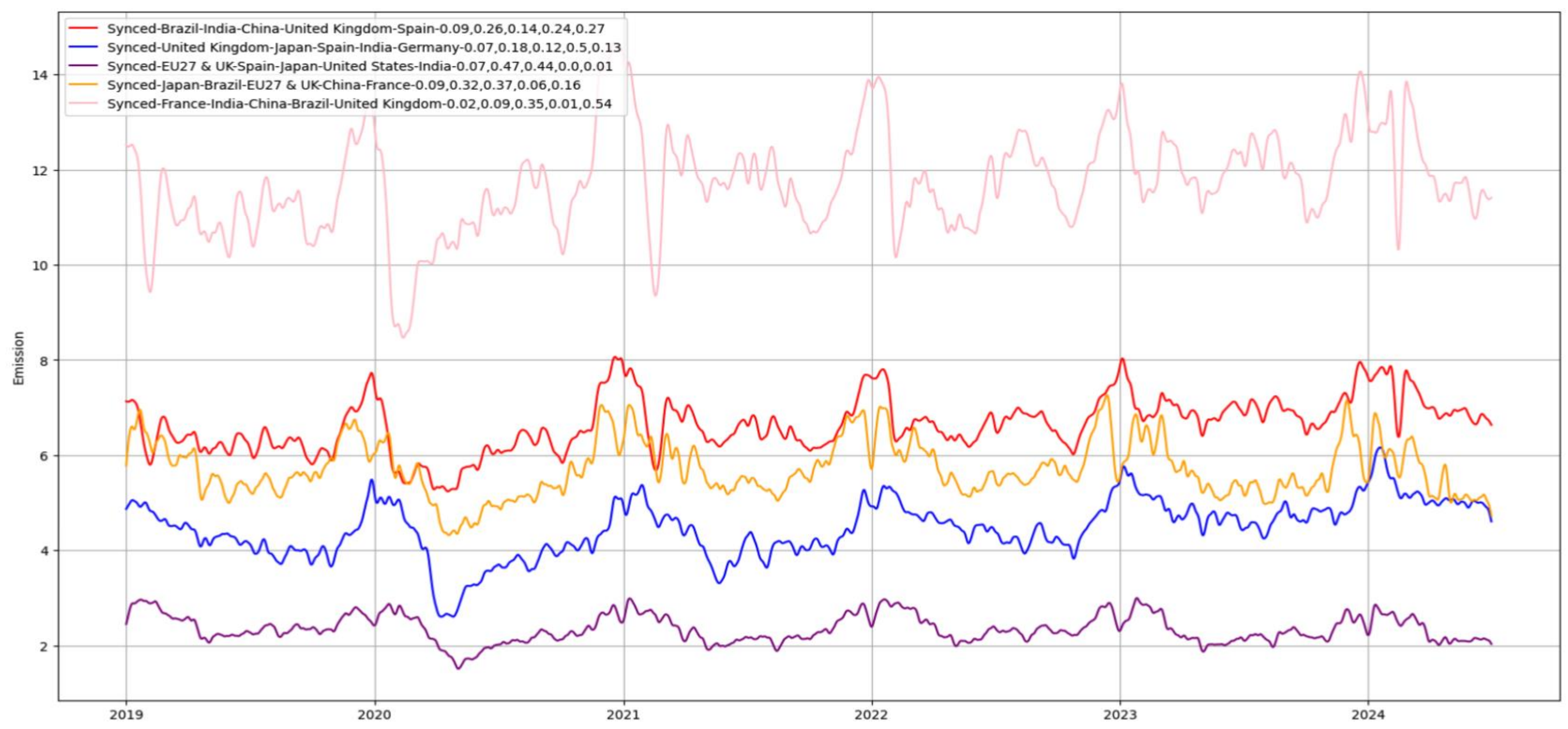
We compare the results of mean absolute percentage error for emission and energy prediction task. Based on n number synthesized countries. We found that, the more synthesized countries generally help to improve the model performance. This is because the lack of diversity of original dataset only contains 12 countries. Each country got complete different energy structures. This result in sparse data samples in the data space, which leads to model to overfit learning discrete features. When the more synthesized countries introduced help model to see more variations, and learns the intermediate continuous pattern.

Carbon Monitor Dataset & Augmentation

In this study, we utilize a real-time high resolution time series emission and energy data set named carbon monitor. This data set covers power generation and daily emission from various regions. It covers 8 source energy, such as gas, nuclear, solar. And 6 categories of emission such as transportation, residential and industry.

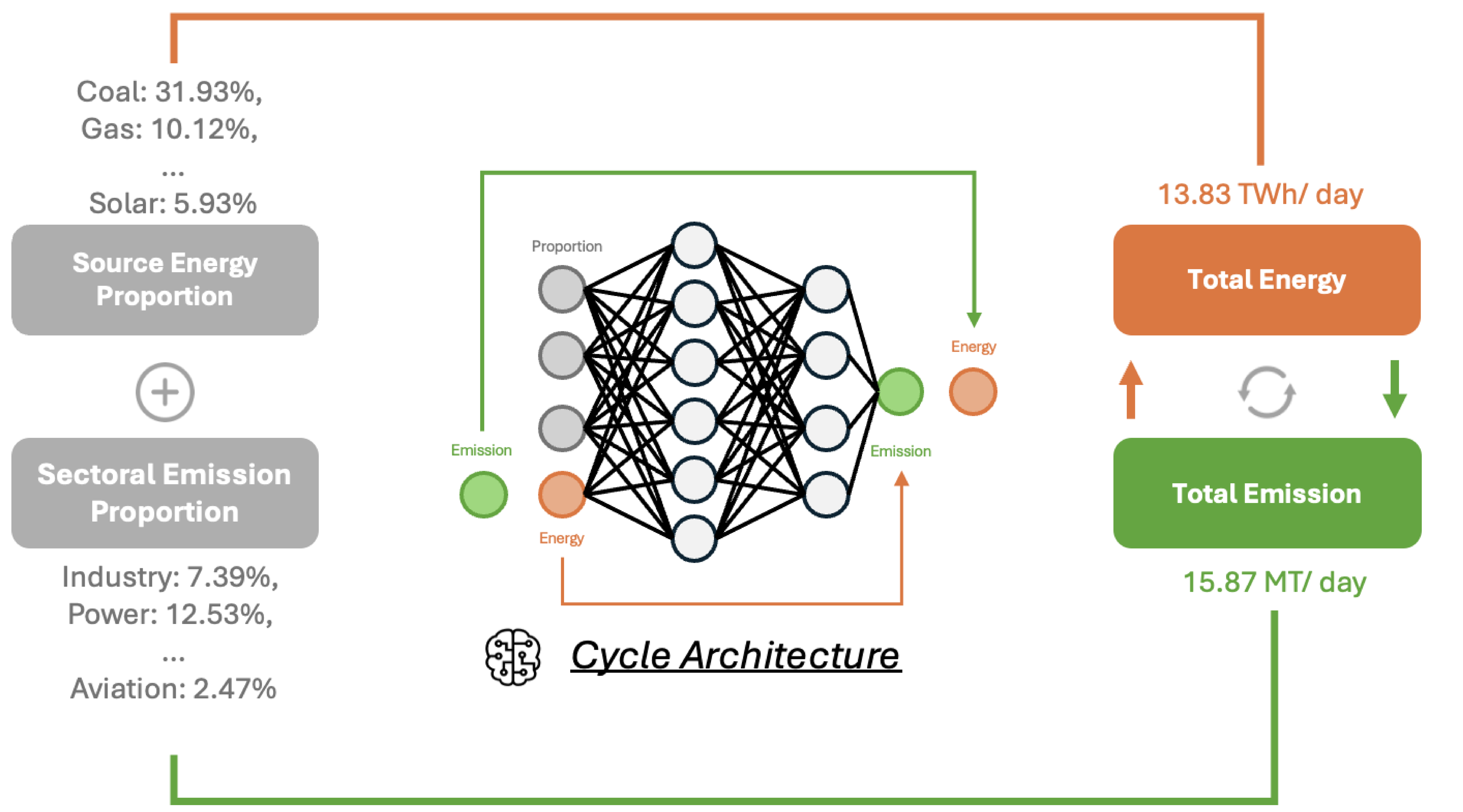






CLEAN: Cycle Learning Energy-Emission Network for Learning Dynamic Mapping

Instead of searching or organizing the data from everywhere to construct a comprehensive emission data. Based on the assumption of relationship between energy and emission. As we observed in right hand side image. We are aiming to build a real-time emission-energy simulator to estimate emission from different source energy such as gas, oil, fossil fuels. Or from one anther size, estimate energy consumption based on emission of different sectors, such as transportation, industry or residence.



Coal	Gas	...	Solar	Total Emission	Total Energy
18.73%	10.25%	...	10.73%	12.88	?
36.61%	8.70%	...	4.42%	?	36.83

Energy Transaction Assessment

Predict the Total Emission or Total Energy while change the Source Energy Proportion

Industry	Power	...	Aviation	Total Emission	Total Energy
12.32%	18.63%	...	5.87%	12.88	?
24.23%	29.67%	...	2.39%	?	36.83

Sector Transaction Assessment

Predict the Total Emission or Total Energy while change the Sectoral Emission Proportion

We formulate the problem into two regression tasks. We normalize the data into proportions as feature. The total energy and total emission are set as label. We predict the total energy based on emission from different sectors, and predict total emission based on mix energy. Both of them got the following application at inference stage. From the emission prediction task, we can perform what-if hypothesis. For example, if we replace 10% of power generation of fossil fuels by gas. What will be the reduced emission? From another size, if we set a industry emission reduction target. How much energy are required to generate? We jointly optimized two tasks by dual encoder architecture at the same time via the proposed cycle loss inspired by CycleGAN helps to learn a consistent feature space serving as the consistency term for regularization.