Design of Experiments (DOE) Execution and Objectives

Hossein Beidaghydizaji

Introduction

This project employs a systematic and statistical Design of Experiments (DOE) methodology to optimize the architecture and hyperparameters of a Convolutional Neural Network (CNN) aimed at gender classification from images. Rather than relying on ad hoc tuning or grid search, DOE allows for efficient exploration of a high-dimensional hyperparameter space, uncovering both main effects and interaction effects with a minimum number of experiments.

The chosen experimental design is the Central Composite Design (CCD) as an approach for modeling complex, nonlinear relationships with relatively few experimental runs. CCD is part of the Response Surface Methodology (RSM) family, making it particularly useful for modeling and optimizing performance metrics that exhibit curvature.

The goal of this DOE is to model the behavior of CNN performance metrics (F1-score, overfitting gap, and trainable parameters) as functions of six key hyperparameters, and to identify optimal regions within this design space.

Why Use CCD for CNN Optimization?

- CNNs have nonlinear and interactive relationships between architecture/hyperparameters and performance.
- CCD captures these interactions using a structured and efficient sampling strategy.
- Compared to grid search or random search, CCD achieves better model coverage using fewer runs.

Execution Pipeline

The 81 configurations, as outlined in the final table of this document, were executed in a random sequence to minimize any potential bias arising from system or data variability. For each configuration:

- The CNN is trained on the gender classification dataset.
- Metrics (F1-score, training/validation loss, number of parameters) among the other outputs are recorded (Please see the uploaded results excel sheet for further information).
- Runs are replicated at center points to provide an internal estimate of experimental error.

- Additionally, all designed experiments were also repeated, and the average responses were employed for further statistical analysis and advanced controlled optimization.

Information about the design of the experiments (DOE) for the optimization of convolutional neural network (CNN) for image classification (gender recognition):

DOE Type	**Central Composite Design (CCD)**
Factors	6 factors (see below)
Responses	Trainable parameters, F1-Score accuracy and overfitting gap
Replicates	Replicates are for reliability check
Randomization	Random run order (to reduce systematic errors)

Factors (hyperparameters) and their levels for optimization:

Factor Name	**Units**	**Low (-1)**	**Center (0)**	**High (+1)**	**Axial (-α)**	**Axial (+α)**
Learning Rate	-	0.0005	0.001	0.005	0.00025	0.01
Dropout Rate	-	0.2	0.3	0.5	0.1	0.6
Number of Conv2D Layers	int	2	3	4	1	5
Units Before Last Layer	int	1024	2048	4096	512	8192
Filters in Layer 1	int	8	16	32	4	64
Filters in Layer 2	int	16	32	64	8	128

With 6 above factors, a CCD full-factorial design is represented as follows:

Item	**Count**
Full factorial points (2^6)	64
Axial points (2 * 6)	12
Center points	5 (for good error estimation)
α value	2^(1/4) ≈ 1.682
Total runs	81 runs

The following table lists all the designed runs for DOE analysis:

Random Run Order	Learning Rate	Dropout	Convolutional Layers	Units of Last Layer	Filters of Layer 1	Filters of Layer 2	Run Type
1	0.005	0.2	4	4096	32	16	Factorial
2	0.0005	0.2	2	1024	8	16	Factorial
3	0.005	0.2	4	1024	32	16	Factorial
4	0.005	0.5	4	4096	32	16	Factorial
5	0.005	0.2	2	1024	32	16	Factorial
6	0.0005	0.2	4	4096	32	16	Factorial
7	0.005	0.2	2	4096	8	16	Factorial
8	0.001	0.3	3	512	16	32	Axial
9	0.0005	0.2	4	1024	8	16	Factorial
10	0.0005	0.2	4	4096	8	16	Factorial
11	0.0005	0.5	2	1024	32	64	Factorial
12	0.0005	0.5	2	1024	8	64	Factorial
13	0.001	0.6	3	2048	16	32	Axial
14	0.005	0.5	2	1024	8	64	Factorial
15	0.001	0.3	1	2048	16	32	Axial
16	0.0005	0.5	4	4096	8	64	Factorial
17	0.001	0.3	3	2048	64	32	Axial
18	0.0005	0.5	4	4096	32	64	Factorial
19	0.005	0.5	4	1024	32	64	Factorial
20	0.0005	0.2	2	4096	8	64	Factorial
21	0.0005	0.5	2	4096	8	16	Factorial
22	0.00025	0.3	3	2048	16	32	Axial
23	0.0005	0.5	4	1024	8	16	Factorial
24	0.005	0.5	4	4096	8	64	Factorial
25	0.005	0.2	2	1024	8	64	Factorial
26	0.005	0.2	4	4096	32	64	Factorial
27	0.005	0.2	2	4096	8	64	Factorial
28	0.005	0.2	4	1024	32	64	Factorial
29	0.0005	0.2	2	1024	32	16	Factorial
30	0.005	0.5	4	1024	8	64	Factorial
31	0.0005	0.2	2	4096	32	64	Factorial
32	0.001	0.3	3	2048	16	32	Center
33	0.005	0.5	4	1024	8	16	Factorial
34	0.005	0.2	2	1024	32	64	Factorial
35	0.0005	0.5	4	1024	32	64	Factorial
36	0.005	0.5	2	1024	32	16	Factorial
37	0.001	0.1	3	2048	16	32	Axial
38	0.0005	0.5	2	4096	32	16	Factorial
39	0.0005	0.2	4	4096	8	64	Factorial
40	0.0005	0.5	4	4096	8	16	Factorial
41	0.001	0.3	3	2048	16	32	Center
42	0.005	0.5	2	1024	8	16	Factorial

43	0.0005	0.5	2	1024	32	16	Factorial
44	0.005	0.2	4	1024	8	64	Factorial
45	0.0005	0.2	2	4096	8	16	Factorial
46	0.01	0.3	3	2048	16	32	Axial
47	0.005	0.2	4	1024	8	16	Factorial
48	0.0005	0.2	4	1024	8	64	Factorial
49	0.001	0.3	3	2048	4	32	Axial
50	0.005	0.2	2	4096	32	64	Factorial
51	0.005	0.2	4	4096	8	64	Factorial
52	0.001	0.3	3	2048	16	32	Center
53	0.005	0.5	4	4096	8	16	Factorial
54	0.005	0.5	2	4096	32	16	Factorial
55	0.0005	0.5	2	4096	8	64	Factorial
56	0.005	0.2	2	4096	32	16	Factorial
57	0.0005	0.2	2	1024	32	64	Factorial
58	0.0005	0.2	2	4096	32	16	Factorial
59	0.005	0.5	2	4096	8	64	Factorial
60	0.001	0.3	3	2048	16	32	Center
61	0.0005	0.5	2	4096	32	64	Factorial
62	0.005	0.5	2	4096	8	16	Factorial
63	0.0005	0.2	2	1024	8	64	Factorial
64	0.001	0.3	3	2048	16	32	Center
65	0.005	0.5	2	4096	32	64	Factorial
66	0.005	0.5	4	4096	32	64	Factorial
67	0.001	0.3	5	2048	16	32	Axial
68	0.0005	0.5	4	1024	8	64	Factorial
69	0.0005	0.5	4	4096	32	16	Factorial
70	0.0005	0.5	2	1024	8	16	Factorial
71	0.0005	0.2	4	1024	32	64	Factorial
72	0.0005	0.5	4	1024	32	16	Factorial
73	0.005	0.2	2	1024	8	16	Factorial
74	0.005	0.5	4	1024	32	16	Factorial
75	0.001	0.3	3	2048	16	128	Axial
76	0.001	0.3	3	2048	16	8	Axial
77	0.0005	0.2	4	1024	32	16	Factorial
78	0.0005	0.2	4	4096	32	64	Factorial
79	0.001	0.3	3	8192	16	32	Axial
80	0.005	0.2	4	4096	8	16	Factorial
81	0.005	0.5	2	1024	32	64	Factorial

Data Analysis Goals

Post-experiment, regression models and response surface plots are used to:

- Identify statistically significant factors.
- Evaluate the main effects and interaction effects.
- Detect curvature and optimal parameter settings.
- Minimize overfitting while maximizing F1-score accuracy.
- Reduce model complexity (fewer parameters).

Model Interpretation

Statistical modeling from the DOE data will involve:

- ANOVA (Analysis of Variance)
- Regression Coefficient Analysis
- Contour Plots and Surface Maps
- Multi-response Optimization (e.g., desirability functions)

Summary and Future Work

This repository provides a complete view of the DOE process applied to CNN optimization for gender classification. Key takeaways:

- The DOE framework ensures a scientific approach to hyperparameter tuning.
- CCD enables efficient identification of nonlinear dependencies and parameter interactions.
- The results propose a final, tuned CNN configuration that balances accuracy and the model size.

Future directions include:

- Validating the final model on external datasets.
- Improving final model accuracy by connecting it to the commonly trained big CNN model (e.g., ResNet50).
- Automating the DOE pipeline with reinforcement learning agents trained on the response surface.

By combining deep learning with classical experimental design DOE, this project aims to make CNN tuning more interpretable, reproducible, and efficient.