

# Automated Assessment of Residual Plots with Computer Vision Models

## Response to reviewers

2026-01-12

I thank the editor and the reviewer for their constructive comments that have improved this paper. I have addressed all comments, colored in purple. In addition, I also include an annotated version of the paper (`diff.pdf`) with the difference between the original and revised versions produced using `latexdiff`.

### Reviewer

#### General Comments

The idea of the paper seems interesting: using computer vision to help identify residual plots which indicates model violation. For any computer vision task, the input is clearly the image. However, there should be a clear definition of the outcome variable that the algorithm wants to predict from images. From the current draft of this paper, it is unclear that what is the outcome variable that the computer vision is learning from the residual plots, how the training datasets are created for computer vision learning, and how the performance of the computer vision algorithm is assessed. Please see my detailed comments in items 10, 12,13. I suggest the authors make these basic setups explicitly defined from the very beginning of the article.

#### Specific Comments

1. p3, l9: What is “the lineup protocol”?
2. missing references: Loy and Hofmann (2013; 2014; 2015)
3. p8 l16: Why do we need to replace it by a full-rank covariance matrix?
4. eq (2), typically KL is specified by  $KL(p — q)$  due to asymmetry

5. p9 l53: to solve eq 2: using “evaluate” for “solve” is better
6. p19 l30: What is “the data generating process”? We don’t know the true distribution of  $y$ .
7. p12, sec 5.1: this sounds like a standard simulation of the sampling distribution of  $\hat{D}$  in traditional statistics.
8. p13, Sec 5.2: How do you do bootstrapping? We need to know the distribution of  $\hat{D}$  under the null. However, the given observed data  $y$  may not come from the null.
9. Tbl1: it is unclear what this table is measuring. What is the R2 measuring? What’s the response and what are the predictors?
10. Section 7 and 8: In computing  $\hat{D}$ , you need a P and Q for each targeted model violation. Is your computer vision learning algorithm targeted a particular model departure, eg, non-linearity or heteroskedasticity? What are your P and Q in generating the training data of  $\hat{D}$  and residual plots for computer vision learning? However, your results also show your performance for different kinds of model violations. This is confusing. You need to specify clearly what is the “true”  $\hat{D}$  and what the “predicted  $\hat{D}$ ” using computer vision in generating training data of  $\hat{D}$  and residual plots.
11. There is no numbering for your equations.