TFM

1. **Me mire el algoritmo y en principio sólo utiliza el output de la parte privada f\_\theta\_A, ¿esto es así?**

Sí, el algoritmo en el archivo [attack.py](vscode-file://vscode-app/c:/Users/HugoPrieto/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html) utiliza el output de la parte privada del modelo, específicamente el output de la capa relu7 del clasificador Classifier. Este output se utiliza para la reconstrucción de las imágenes en el modelo de inversión Conv\_4x4\_8x32.

Aquí está el fragmento relevante del código:

with torch.no\_grad():

    prediction = classifier(data, relu=7)

reconstruction = inversion(prediction)

En este fragmento, el output de la capa relu7 del clasificador se pasa al modelo de inversión para reconstruir las imágenes.

1. **Arquitectura del modelo de inversión G en la variante white-box. En principio, es una transposed NN, qué forma tiene.**
2. **En algoritmo de entrenamiento de G, se menciona esto: Randomly sample 𝑓𝜃𝐴 (𝑥1), 𝑓𝜃𝐴(𝑥2), . . . , 𝑓𝜃𝐴(𝑥𝑘 ) from F. Esto es, como se obtienen las muestras para entrenar G.**
3. **Y verificar que sólo se utiliza el output de la parte privada 𝑓𝜃𝐴, esto es, le da igual el resto de capas de la parte pública.**