

Índice de contenidos

Índice de contenidos	V
Resumen	VII
Abstract	IX
1. Introducción	1
2. Marco teórico	5
3. Generación de ejemplos adversos	13
4. Detección de ejemplos adversos	17
5. Resultados	21
5.1. Bases de datos	21
5.2. Arquitecturas de red	21
5.3. Ataques	25
5.4. Similitud con la matriz de confusión	26
5.5. Un enfoque de detección basado en la relación señal ruido	27
5.6. Método de detección estocástico	30
5.7. Comparación con otros trabajos	33
6. Conclusiones	35
A. Histogramas de robustez	39
B. Curvas ROC-AUC	41
Bibliografía	43