الإستهلال

قال تعالى :

(نَرْفَعُ دَرَجَاتٍ مَنْ نَشَاءُ وَفَوْقَ كُلِّ ذِيعِلْمٍ عَلِيم)

صدق الله العظيم

(سورة يوسف الاية : 76)

Dedication

This work is dedicated to Our parents; who has always been the beacon through darkest nights.

Our brothers and sisters who support us in each way possible

Ali Malik, Salma Elyas, Hussein Eldaw, abdelhamed, T. Hanaa Mohammed , T. Malaz Omer and Minas for supporting us all the way through

And everyone who supported us during this work without whom none of this work would be done.

ACKNOWLEDGEMENT

Firstly, we do acknowledge the big role our supervisor Dr. Eiman Omer did, for his support in the project and for the guidance offered to provide all required knowledge faithfully. She was keen on the interview and the weekly supervision so that he was fully aware of how the project was going, so that she was choosing the suitable time for us to experiment so as not to waste the effort in repeating it more than once. She also took care of the tiniest details in order to complete the project to the fullest. And we are grateful to our colleagues in electronic engineering students (Batch 19), Last but not least we would like to thank our families; for their immense support and their great love that lead us to where we are now.

ABSTRACT

Artificial intelligence (AI) and machine learning. Algorithms receive input data and use statistical analysis to predict the outcome, thus giving the ability to the ability to think like humans in a way that helps us to use it in different applications in daily life like self-driving cars, spam detectors, machine-learning powered scanner scans suitcases for weapons at the airport. But unfortunately these algorithms, despite their high intelligence, can be tricked into making mistakes using the adversarial attack, there are several known methods for crafting adversarial examples, and they vary greatly with respect to complexity, computational cost, and the level of access required on the attacked model. pretrained image classification models have been used to perform the attack on. The first is LeNet with 74.8% top-1 accuracy and the second ResNet with 92.3% top-1 accuracy. Trained on standard datasets (MNIST and CIFAR-10). the two methods have compared on: The average distortion to the original image. Time and computing resources it takes to perform the attack. The percentages of getting a successful attack on the first attempt. The resistance of the models was attacked against each method of attacks. and also noticed that in the non-targeted attack in each of the two methods, the incidence rate is very high, and the higher the value of the epsilon or the greater number of targeted pixels, the attack occurs in a shorter time.

المستخلص

الذكاء الاصطناعي وتعليم الالة عبارة عن خوارزميات تستقبل مدخلات وتقوم بتوقع النتيجة , مما يمنح الالة القدرة على التفكير مثل البشر بطريقة تساعدنا على استخدامها في حياتنا اليومية مثل السيارات ذاتية القيادة , جهاز كشف البريد العشوائي وماسحات الحقائب عند المطارات ولكن للاسف هذه الخوارزميات رغم ذكائها يمكن خداعها باستخدام الهجوم المعادي . هنالك الكثير من الطرق المعروفة لعمل هذا الهجوم وتتفاوت حسب تعقيدها والتكاليف الحسابية ومستوى الوصول المطلوب الى الخوارزمية المستهدفة , قمنا باستخدام خوارزميات تصنيف تم تدريبها مسبقاً لتنفيذ الهجوم عليها , الاول هوLeNet بدقة تبلغ 74.8% والثاني ResNet بدقة تبلغ 92.3% تم تدريبهم على مجموعة من البيانات القياسية (CIFAR-10 , MNIST) ,قمنا باستخدام google colab كمنصة برمجة و تطوير ولغة بايثون لتنفيذ وتصميم خوارزمية الهجوم , قمنا بمقارنة طريقتين من طرق الهجوم من عدة نواحي : متوسط التشوه المضاف للصورة الاصلية , الوقت وموارد الحاسوب التي يستغرقها تنفيذ الهجوم , نسبة النجاح في الهجوم من المحاولة الاولى , مقاومة الخوارزميات المتسهدفة لطرق الهجوم المستخدمة .

List of Content