العلامة المائية الرقمية للشبكات العصبونية العميقة

كلية الهندسة – قسم الحاسوب

إشراف: د. سمر الحلبي

إعداد: الطالب فريد شيخ الشباب

الملخص:

شهدت الآونة الأخيرة تقدماً واسعاً للشبكات العصبية العميقة التي قادنا إلى تطبيقات قوية في عدة مجالات كمعالجة الصور، التعرف على الكلام، معالجة اللغات الطبيعية وغيرها، كما سهلت نماذج الشبكات العصبية العميقة المدربة على الباحثين القيام بالعديد من المهام الصعبة، كل ذلك جعل أهمية حماية الملكية الفكرية للشبكات العصبية العميقة حاجة ضرورية وقد كانت العلامة المائية الرقمية إحدى الطرق المحققة لذلك.

سنعرض في هذه المقالة مفهوم العلامة المائية الرقمية للشبكات العصبية العميقة والمتطلبات الأساسية التي يجب أن تحققها، كما سنبين طرق تصنيف العلامة المائية الرقمية بالاعتماد على طرق تضمينها وطرق استخراجها، وما هو الفرق بين العلامة المائية متعددة البت والعلامة المائية صفر بت، وأخيرا سنذكر بعض المعلومات المستخدمة لاكتشاف العلامة المائية الرقمية ومحاولة ازالتها.

المقدمة:

تعدُّ الانترنت البيئة الأوسع للتعامل مع البيانات في ثورة المعلومات، ولازدياد أهميتها برز التفكير الجاد في حمايتها وحماية خصوصية الأشخاص المستخدمين لها، فلم يعد موضوع الأمن متعلقاً بالبحث في ثغرات بروتوكولات الاتصالات فحسب بل أصبح يشمل أيضاً حماية محتوى البيانات المتداولة عبرها.

يتم نشر نماذج التعلم العميق واستخدامها ومشاركتها على نطاق واسع هذه الأيام، وعملية تدريب نموذج التعلم العميق مهمة هامة وتتطلب كميات هائلة من البيانات الخاصة، وتستهلك قدرًا هائلاً من موارد الحاسوب والطاقة والخبرة البشرية، ولكن ماذا يمنع لو أراد بعض الأشخاص سرقة النماذج ونسبها لهم.

لهذا السبب يجب حماية النموذج بتقنيات حماية الملكية الفكرية، ومنها تقنية العلامة المائية الرقمية (Digital Watermark (WM))، أي يمكننا تضمين العلامات المائية الرقمية في الشبكات العصبية العميقة ((Deep Neural Network (DNN).

العلامة المائية:

أتى مصطلح العلامة المائية من أصل ألماني أطلق على تأثير الماء في نوع معين من الورق، وظهرت العلامات المائية الورقية في فن الصناعة اليدوية حيث وجدت أقدم ورقة بعلامة مائية في الأرشيف في سنة 1292 في مدينة، وقد لعبت دوراً رئيساً في تطور الصناعات الورقية، حيث كانت الطريقة المثالية للقضاء على أي احتمالية للتشويش في تمييز إنتاج المصانع وإظهار العلامات التجارية.

ظهرت أول تقنية مشابهة للعلامة المائية الرقمية في عام 1945 من قبل Emil Hembrooke لتمييز الأعمال الموسيقية، ولا تزال العلامة المائية الورقية تستخدم في السندات المالية والعملات الورقية حتى وقتنا الحاضر لمنع تزويرها .

تُعرف العلامة المائية الرقمية بأنها إضافة بيانات معينة متعلقة بالملف الأصلي المراد تضمين العلامة المائية فيه أو متعلقة بمالك هذا الملف (رقم، الاسم، التوقيع، بصمة الإصبع، ...) إلى وسط حامل لهذه العلامة (ملفات الصوت – الفيديو – الصور– الشبكات العصبية–...) بهدف الاستفادة من هذه البيانات فيما بعد في عدة أمور، ويشترط عدم تأثير هذه الإضافات على الملف الأصلي، بالإضافة إلى أنها لا تعيق استخدام بيانات الملف الأصلي بل تزودها بآلية لمنع تزييفها.

اكتسبت العلامة المائية الرقمية أهميتها لمساهمتها في الحفاظ على حقوق الطبع والنشر والتأليف والملكية في ظل تزايد عمليات القرصنة والاستنساخ غير المشروع خاصة عبر الإنترنت.

العلامة المائية الرقمية في الشبكات العصبية العميقة [2][3]:

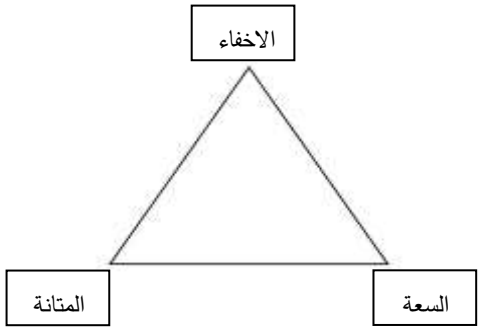
يتم إضافة العلامة المائية الرقمية بشكل عام إلى إشارة الملف –الوسط الحامل للعلامة المائية الرقمية– ضمن البتات الأقل أهمية أو بشكل عشوائي مع الحفاظ على عدم التأثير بشكل واضح على إشارة الملف الأصلية.

أما العلامة المائية الرقمية في الشبكات العصبية العمية فإننا نختار إضافتها في مرحلة التدريب إلى أوزان الشبكة مع المحافظة على عدم التأثير على قدرة الشبكة من أداء المهمة الأساسية المصممة لها.

كما يستخدم مفتاح لتشفير العلامة المائية الرقمية لإضفاء السرية على عملية وضعها وجعل عملية كشفها واستردادها أمر صعب للمستخدمين غير المصرح لهم بذلك، أي أن مالك المفتاح وحده يستطيع كشف العلامة المائية واظهارها.

متطلبات العلامة المائية الرقمية في الشبكات العصبية العميقة [2]:

تظهر متطلبات مختلفة ومتعددة تبعاً للتطبيق الذي يستخدم به العلامة المائية الرقمية والهدف منه، ولكن تم تلخيص أهم متطلبات العلامة المائية الرقمية في المثلث أدناه (الشكل 2):



الشكل (2) متطلبات العلامة المائية الرقمية في DNN

1. **الاخفاء:** وهي من أكثر المتطلبات أهمية وتعني عدم رؤية العلامة المائية الرقمية وأن التعديلات على معلومات المحتوى الأصلي للشبكة العصبية العميقة يبقى دون العتبة المحسوسة.
2. **السعة:** ويقصد بها عدد بتات(Bits) العلامة المائية الرقمية المشفرة التي يمكن إضافتها إلى الشبكة العصبية العميقة.
3. **المتانة:** تعني القدرة على استخراج العلامة المائية الرقمية بشكل صحيح حتى عندما يتم تعديل نموذج الشبكة العصبية، وعمليتي التلاعب الأكثر شيوعًا التي يجب على العلامة المائية الرقمية في DNN مقاومتهما:
   a. **الضبط الدقيق (Fine–tuning):** وهي عملية إعادة تدريب الشبكة العصبية العميقة لحل مهمة جديدة بعد أن تم تدريبه في البداية على حل مهمة معينة، والذي قد يؤدي إلى تغيير أوزان نموذج العلامة المائية، لذلك من المهم التأكد من أن العلامة المائية الرقمية مقاومة لعمية الضبط.

b. **تقليم الشبكة (Network pruning):** تعني عملية التقليم في الشبكة العصبية العميقة هي قطع الأوزان التي تقل قيمتها المطلقة عن الحد الأدنى للعتبة التي تم وضعها للشبكة. يتم تطبيق تقليم الشبكة لتبسيط نموذج شبكة عصبية معقد لإمكانية استخدامه على الأجهزة ذات الإمكانيات المتوسطة. نظرًا لتغيير الأوزان في عملية التقليم فمن الضروري أن تكون العلامة المائية الرقمية مقاومة لذلك.

4. متطلبات إضافية للعلامة المائية الرقمية:

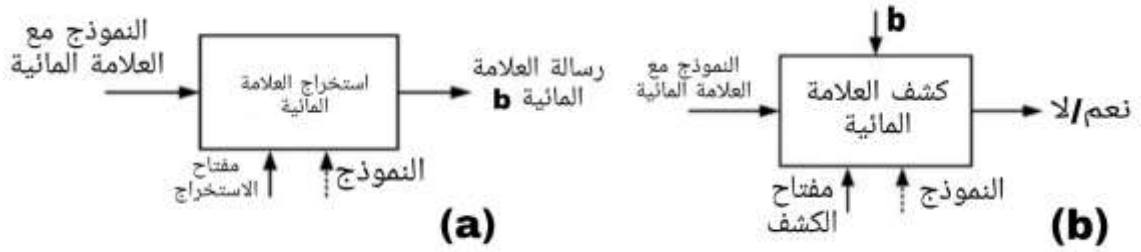
i. **الأمن (Security):** لا يمكن فقدان العلامة المائية الرقمية إلا من خلال التدهور الكبير في نموذج DNN المضيف، وقد ركزت الأبحاث بشكل أساسي على نوع من الهجمات ألا هو الكتابة فوق العلامة المائية، أي محاولة إضافة علامة مائية رقمية إضافية إلى النموذج لجعل العلامة المائية الأصلية غير ملحوظة.

ii. **العمومية (Generality):** يجب أن تكون خوارزميات العلامات المائية الرقمية لشبكات العصبية العميقة قابلة للتكيف مع مجموعة من المعماريات المختلفة للشبكات.

نماذج العلامات المائية الرقمية في الشبكات العصبية العميقة:

تصنف نماذج العلامة المائية الرقمية فيDNN إلى:

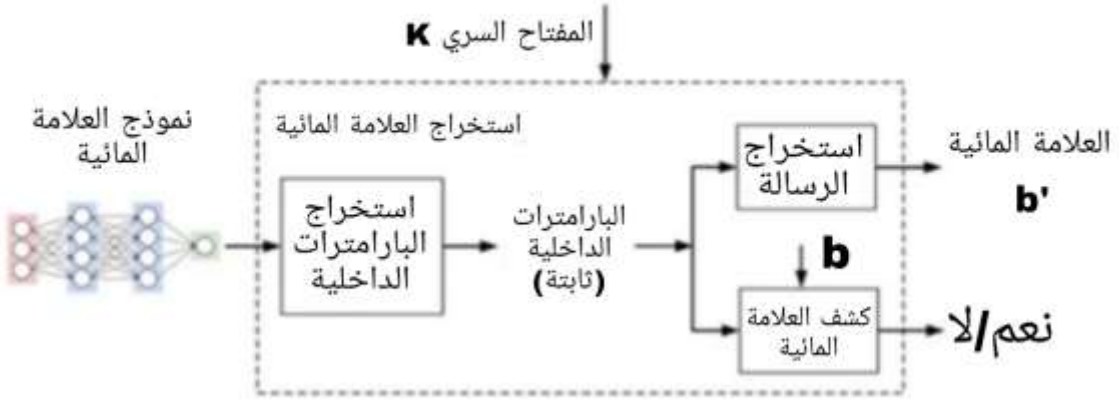
1. **العلامة المائية متعددة البت والعلامة المائية صفر بت [2]:** يمكن التمييز بين فئتين من خوارزميات العلامة المائية الرقمية اعتمادًا على المحتوى الدقيق للعلامة المائية المستخرجة هما: العلامة المائية متعددة البت والعلامة المائية صفر بت. حيث عند استخراج العلامة المائية متعددة البتات يطلب استخراج العلامة المائية المؤلفة من تسلسل N بت، بينما عند استخراج العلامة المائية صفر بت فيطلب كشف وجود علامة مائية أم لا. (الشكل 3)



الشكل (3) يوضح a العلامة المائية الرقمية متعددة البت، b يوضح بكشف العلامة المائية الرقمية صفر بت

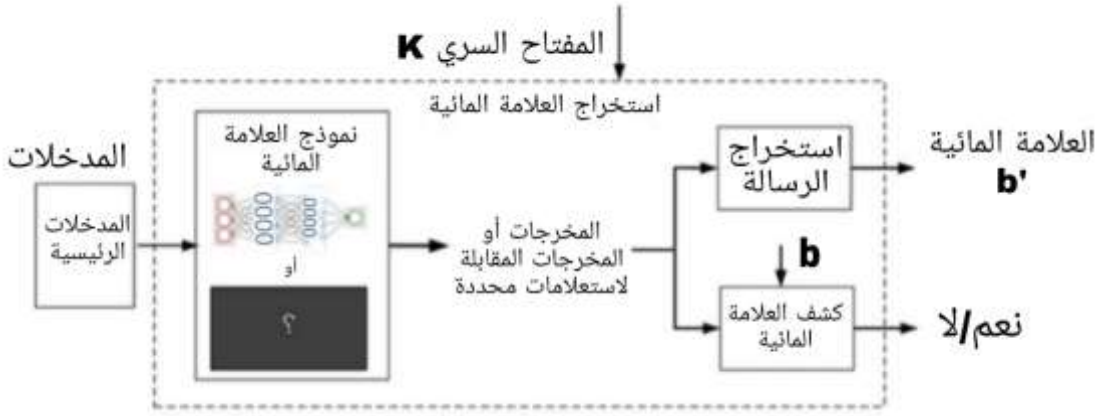
2. **العلامات المائية الثابتة والعلامات المائية الديناميكية [2]:** تصنيف العلامة المائية لـ DNN إلى علامة مائية ثابتة وعلامة مائية ديناميكية اعتمادًا على المكان الذي يمكن قراءة العلامة المائية منه.

a. **العلامة المائية الثابتة** [5][4]: يمكن قراءة العلامة المائية مباشرة من أوزان ثابتة تم تحديدها في مرحلة التدريب ولا تعتمد على مدخلات الشبكة العصبية، وتعتبر مشابهة لتقنيات العلامات المائية التقليدية للوسائط المتعددة. (الشكل 4)



الشكل (4) يوضح كيفية تضمين العلامة المائية الثابتة

b. **العلامة المائية الديناميكية [6]** : يتم تضمين العلامة المائية من خلال أوزان مختارة بشكل دقيق من الشبكة حسب مدخلات محددة سابقاً تدعى بمدخلات التشغيل أو المدخلات الرئيسية.

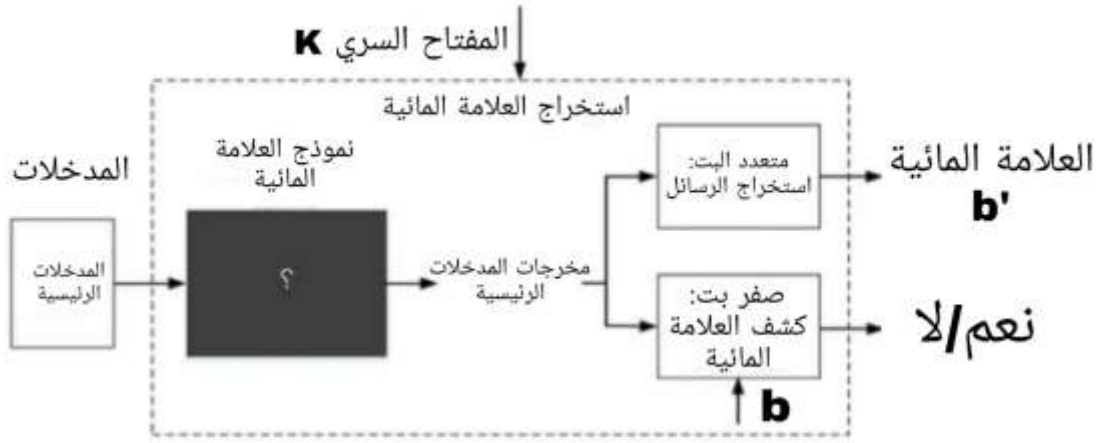


الشكل (5) يوضح كيفية تضمين العلامة المائية الديناميكية

استخراج العلامة المائية الرقمية من الشبكات العصبية العميقة [2][3]:

تقسم عملية استخراج العلامة المائية من DNN استنادًا إلى البيانات التي يمكن لمستخرج العلامة المائية الوصول إليها إلى طريقة الصندوق الأبيض وطريقة الصندوق الأسود.

1. **طريقة الصندوق الأبيض:** في هذه الطريقة يجب أن تتوفر الأوزان الداخلية لنماذج DNN، ويمكن أن تكون علامة مائية ثابتة أو ديناميكية.
2. **طريقة الصندوق الأسود:** في هذه الطريقة يمكن الوصول إلى المخرج النهائي لنموذج DNN، أي الشيء الوحيد الذي يمكن معرفته هو مدخلات النموذج والمخرجات الناتجة عنها، وبالتالي يمكن استخراج العلامة المائية عن طريق مقارنة مخرجات النموذج بمجموعة من المدخلات المختارة بشكل دقيق، ولا يمكن استخدام هذه الطريقة إلا من خلال العلامة المائية الديناميكية. (الشكل 6)



الشكل (6) يوضح استخراج العلامة المائية باستخدام طريقة الصندوق الأسود

الهجمات على العلامة المائية للشبكات العصبية العميقة [1][2][3]:

اقترحت عدة مجموعات بحثية هجمات تحدد وجود علامة مائية ضمن أوزان الشبكة العصبية العميقة وطرق لإزالتها منها. فمثلاً يمكن للمهاجمين الاستفادة من أن تضمين العلامة المائية يزيد من تباين الأوزان في الشبكة، وبالتالي يجعل من الممكن التمييز بين الشبكة التي تتضمن علامة مائية والشبكة التي لا تتضمن علامة مائية. كما أن أيضاً الانحراف المعياري لأوزان الشبكة يزداد خطيًا مع بُعد العلامة المائية، وهذا يسمح للمهاجم بتقدير طول العلامة المائية أو معرفة ما إذا كانت موجودة أم لا. يمكن للمهاجمين بعد التأكد من المعلومات السابقة استبدال العلامة المائية الموجودة بعلامة جديدة، مما يجعل العلامة المائية الأصلية غير قابلة للقراءة.

الخاتمة:

أصبحت الشبكات العصبية العميقة ذات شعبية متزايدة بفضل قدراتها الشبيهة بقدرات الإنسان، وبالنظر إلى الموارد المستثمرة في صنعها فمن المهم حماية هذه التطورات وحماية الملكية الفكرية لها، وقد كانت العلامة المائية الرقمية هي إحدى الطرق الموثوقة لتحقيق ذلك هدف. مما جعل الباحثين يقوموا بتطوير فئات جديدة من الخوارزميات لتضمن العلامات المائية الرقمية في نماذج الشبكات العصبية العميقة ومحاولة توفير المتانة ضد الضبط الدقيق وتقليم النماذج وهجمات إعادة الكتابة.

المراجع:

[1] Jae-Eun Lee, Young-Ho Seo, Dong-Wook Kim, 2020. "Convolutional Neural Network-Based Digital Image Watermarking Adaptive to the Resolution of Image and Watermark", www.mdpi.com/journal/applsci.

[2] Yue Li, Hongxia Wang, Mauro Barni, 2021. "A survey of deep neural network watermarking techniques".

[3] Farah Deeba, She Kun, Fayaz Ali Dharejo, Hameer Langah, Hira Memon, 2020. "Digital Watermarking Using Deep Neural Network", International Journal of Machine Learning and Computing, Vol. 10, No.2.

[4] Uchida, Y., Nagai, Y., Sakazawa, S., Satoh, S., 2017. "Embedding watermarks into deep neural networks", in: Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, pp. 269_277.

[5] Li, Y., Tondi, B., Barni, M., 2020. "Spread-transform dither modulation watermarking of deep neural network". arXiv preprint arXiv:2012.14171.

[6] Rouhani, B.D., Chen, H., Koushanfar, F., 2019. "Deepsigns: an end-to-end watermarking framework for protecting the ownership of deep neural networks", in: The 24th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS). ACM.