

SpeakLiz من تالوف: نحو تطبيق جوال للتعرف على لغة الإشارة

بدعم من برنامج مدى للابتكار

هوجو جاكوم أندراي1، شهباز أحمد 2
تالوف1، مركز مدى 2

ملخص:

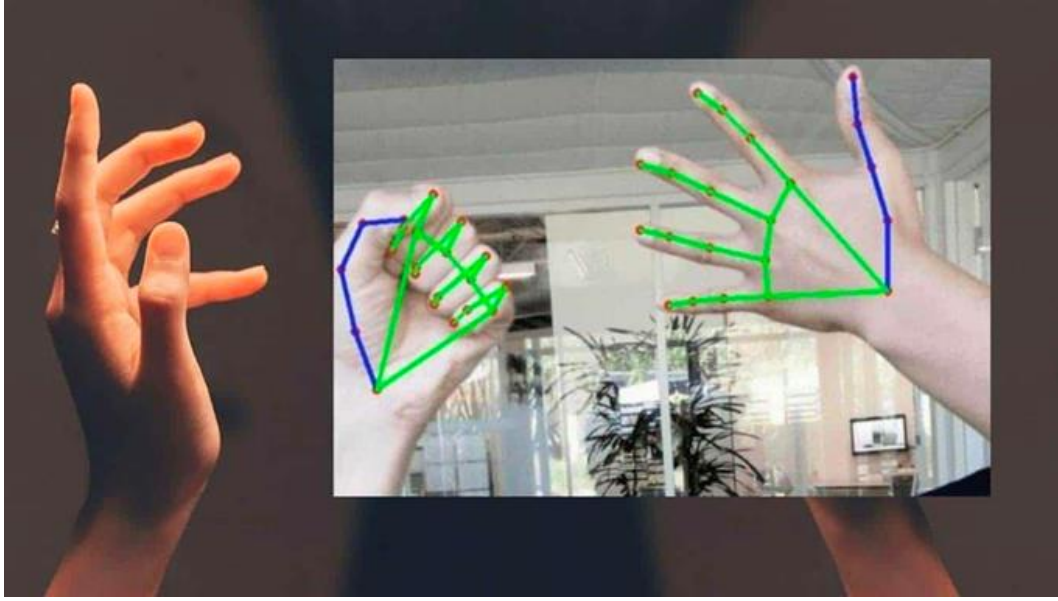
يناقش هذا المقال تكنولوجيا التعرف على لغة الإشارة والتي تعد جزءًا من واحدة من أكثر خوارزميات الذكاء الاصطناعي (AI) تحديًا: التعرف على الإيماءات باستخدام الكاميرا. إنه تطبيق عملي جديد نسبيًا بالمقارنة مع وجوده القديم والنظري في مجال الذكاء الاصطناعي العام. وربما يرجع ذلك إلى قوة الحوسبة المطلوبة لجهاز ما لتتبع الإيماءات المعقدة (مثل لغات الإشارة) فقط باستخدام كاميرا جنبًا إلى جنب مع ظروف الإضاءة والتركيز وثبات حركة الكاميرا غير المثالية أو حتى العديد منها مجتمعة. ومع ذلك، فإن الوصول إلى نتائج مقبولة على الأجهزة المحمولة بمهمة التعرف هذه يعد أمرًا بعيد المنال إذا ما قورن بالقيام بذلك على أجهزة الكمبيوتر (في الواقع، معظم المشاريع البحثية في هذا المجال تعتمد على الكمبيوتر).
الكلمات الرئيسية: التعرف على لغة الإشارة.

المقدمة

يبدأ كل شيء في الذكاء الاصطناعي بمجموعة البيانات (صورة أو نص أو صوت أو مستند). وبغض النظر عن التطورات العديدة في موثوقية الذكاء الاصطناعي مع كميات أقل وأقل من البيانات، لا يزال العثور على ذكاء اصطناعي يمكن أن يقترب من القدرة البشرية النموذجية للتعلم يعتبر أمرًا بعيد المنال. وهذا يعني أن مجموعات البيانات الخاصة بتغذية أداة التعرف على الإيماءات هذه يجب أن تأخذ في الاعتبار كمية كبيرة من البيانات مع العديد من المتغيرات (اختلافات طفيفة في أداء الإيماءة من شخص إلى آخر والسرعة) [1]. كما أن لغات الإشارة معقدة للغاية في تنوعها عبر مختلف البلدان والمناطق [2]. ويتم توصيل لمعلومات عبر حركات الأصابع البسيطة وإيماءات الوجه؛ لذلك، تصبح عملية التنظيم التفصيلية للغاية داخل مجموعة البيانات ضرورية للحصول على أفضل مدخلات ممكنة لحلول الذكاء الاصطناعي اللاحقة التي ستعالج تلك البيانات.

فهم عملية التعرف على لغة الإشارة

بادئ ذي بدء، يتطلب فهم التعرف على لغة الإشارة من الجانب التكنولوجي فهماً عميقاً لبعض اتجاهات الحوسبة والذكاء الاصطناعي [3]. وتعتبر هذه المهمة معقدة جداً على الأجهزة المحمولة بسبب الوقت المطلوب للمعالجة المحلية والاتصال بالإنترنت مع خادم يقوم بمعالجة البيانات ويرسل تعليقات توضيحية أو مخرجات مماثلة إلى الهاتف المحمول [4]. وفي هذه الحالة، يجب أن يكون مستوى تحسين الموارد الحسابية كبيراً للحصول على أداة مفيدة تأخذ في الاعتبار أفضل تجربة مستخدم ممكنة (UX). وبخلاف ذلك، يمكن للمطورين ببساطة تطوير حل قادر على العمل، على الرغم من أنه سيحرق نظام الهاتف في الشريحة ويسنزف بطاريته في غضون ثوانٍ، مما يجعله عديم الفائدة للتطبيقات الواقعية.



الشكل 1: خوارزمية التعرف على لغة الإشارة (Kumar, A, 2016)

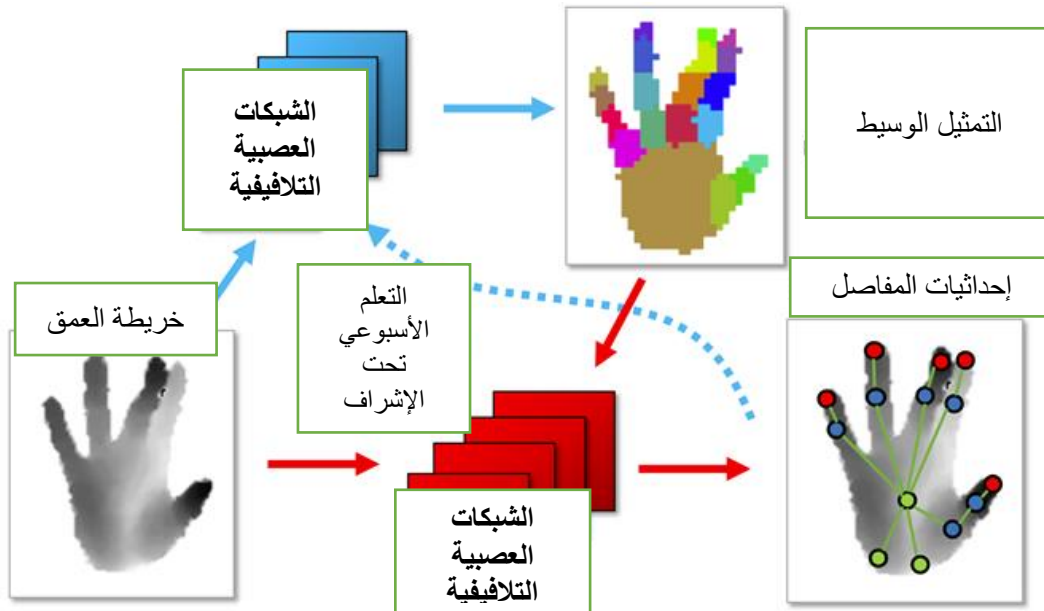
إن الذكاء الاصطناعي هو نظام شامل يغطي جميع مجالات منح الآلات القدرة على أن تكون ذكية [5]. وهنا تتوافق الشبكات العصبية التلافيفية (CNN) مع إحدى الطرق الأكثر شيوعًا لتعلم الآلة لأنها تهدف إلى أداء بعض المهام لمحاكاة كيفية تحديد الشبكات العصبية للدماغ البشري للأنماط وتصنيف المعلومات. وعندما يتعلق الأمر بالتعرف على لغة الإشارة، فإن هناك الكثير من التقنيات ذات الأساليب المختلفة. ومع ذلك، فإن بعضًا من أكثرها شيوعًا هو اكتشاف الأجسام وتحليل الفيديو اعتمادًا على نوع البيانات المراد تحليلها (الإيماءات الثابتة أو تلك التي تتضمن الحركات المعقدة).

لذلك، كان من الضروري التفكير في نوع الشبكات العصبية التلافيفية الذي يمكن أن يعمل بشكل أفضل عندما يكون الهدف هو تحليل أو تتبع البيانات من حركة الإنسان. وهنا يمكن أن يكون البحث عن الشبكات العصبية التلافيفية المصممة للتعامل مع البيانات التي تتغير بطريقة ما بمرور الوقت اختيارًا ممتازًا. ويمكن أن يكون هناك العديد من الخيارات مثل شبكات الذاكرة قصيرة المدى الطويلة (LSTM) [6] والشبكات العصبية المتكررة التلافيفية (SCRNN) من بين أمور أخرى. كما أنه من الضروري ملاحظة أن قرار اختيار الشبكة العصبية يعتمد على نوع مجموعة البيانات المستخدمة في المشروع. ومن الضروري تحليل عدة أمور منذ بداية المشروع بحيث يمكن للباحثين والمطورين تجنب العديد من المفاهيم الخاطئة وتفادي التسبب في تأخير الوقت أو إنفاق الموارد غير الضرورية.

ومع ذلك، فإن الجزء المهم الآخر من هذه الرحلة هو توفير تجربة المستخدم الصحيحة والمحددة للمستخدمين ذوي الإعاقات السمعية، والذي يتضمن العمل على موثوقية الأداة ليس فقط من الجانب التكنولوجي ولكن أيضًا من جانب فائدتها. فمن المعروف أن العديد من أدوات الذكاء الاصطناعي توفر مخرجات رديئة من حيث التعقيد النحوي، ولا يمكنها تجاوز سوى الكلمات الفردية. وبالتالي يصبح تطبيق المعالجة الصحيحة للغة الطبيعية (NLP) أمرًا ضروريًا لسد فجوات التعقيد النحوية بين الجانبين. كما قد تختلف الهياكل النحوية للغات الإشارة عن اللغات المنطوقة. وهذا هو السبب في أن معالجة اللغة الطبيعية

تعد جزءًا حيويًا من هذا اللغز (ليس فقط من حيث التعرف على لغة الإشارة ولكن أيضًا من جهة توليد لغة الإشارة عبر الشخصيات الافتراضية أو أي خيار مشابه) [7]. ويهدف هذا النوع من التكنولوجيا إلى إتاحة هذه التكنولوجيا للجميع وجعل إمكانية الحصول على مساعد رقمي أرخص لفهم البيئة المحيطة في مهام الحياة اليومية بشكل أفضل.

لا يستطيع الجميع تحمل تكاليف خدمات ترجمة لغة الإشارة لأكثر من ساعتين في أفضل الحالات. حتى بالنظر إلى حالة الموارد غير المحدودة لدفع ثمن هذه الخدمات بشكل دائم، لا تزال هناك مشكلة كبيرة وهي الخصوصية [8]. هناك الكثير من اللحظات التي يجب فيها استخدام لغة الإشارة والشخص الذي لا يعرف لغة الإشارة التواصل بينهما فقط دون حضور أي شخص آخر. تلك هي اللحظة التي يصبح فيها هذا النوع من التكنولوجيا ضروريًا للغاية. ويمكن لبقية الحالات، مثل المؤتمرات والبرث التلفزيوني وغير ذلك أن تستمر في طلب خدمات الترجمة البشرية للغة الإشارة.



الشكل 2: معالجة الشبكات العصبية التلافيفية (CNN) (Maro, J, 2020)

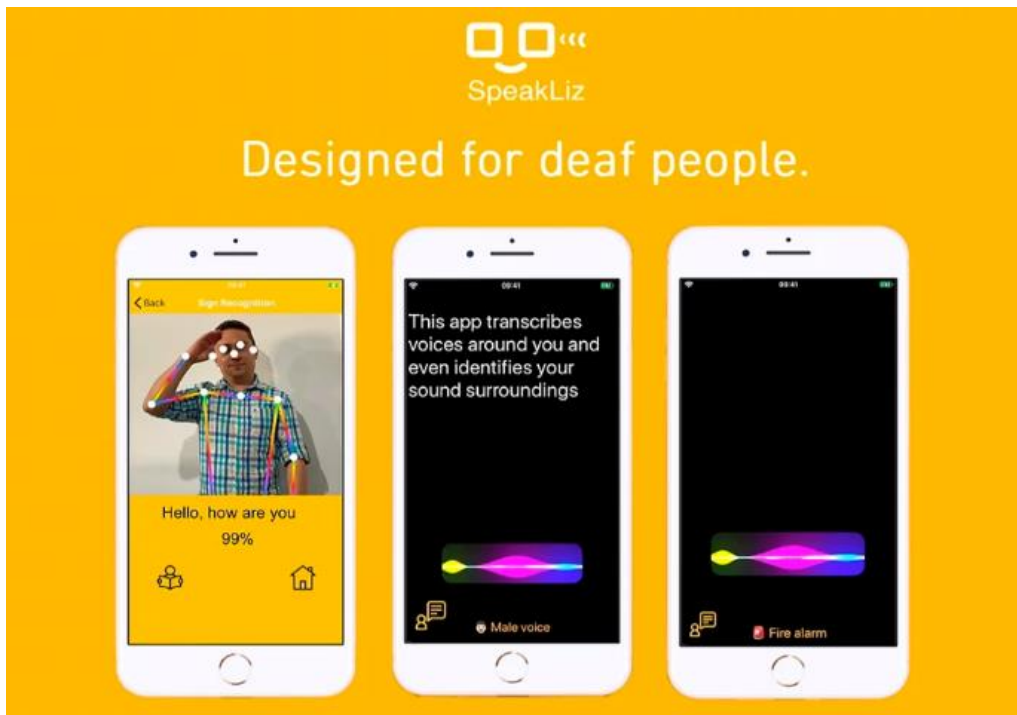
SpeakLiz من تالوف: بدعم من برنامج مدى للابتكار

باختصار، يبدو أن هذه المقاربة للتعرف على لغة الإشارة تمثل تحديًا كبيرًا، ولا يمكن إنكار ذلك، ولكن في الوقت نفسه، يعد رد الفعل الإيجابي لدى المستخدمين أثناء اختبار هذه الأداة دافعًا كبيرًا لمواصلة تطوير مهمة شركة تالوف (التي شارك في تأسيسها كارلوس أوباندو وهوجو جاكوم)، مما يجعل العالم أقرب إلى الجميع بطريقة قابلة للنفاذ مع استخدام أحدث الحلول التكنولوجية القابلة للحمل بسهولة من قبل جميع المستخدمين [9]. ويعد تطوير هذا النوع من الذكاء الاصطناعي لبناء أداة التعرف على لغة الإشارة المحمولة أمرًا صعبًا لأن بناء هذا النوع من التكنولوجيا يتطلب إنشاء العديد من الأشياء من الصفر بما في

ذلك مجموعات البيانات. إن الهدف هو تحديث SpeakLiz (المصمم للأشخاص ذوي الإعاقة السمعية) وتزويده بميزات أفضل، خاصة للتعرف على لغة الإشارة بواسطة الكاميرا.

يحتوي تطبيق SpeakLiz من تأليف على ميزات مثل ميزة التعرف على لغة الإشارة، والتي تحول لغة الإشارة الكلاسيكية إلى شكل حديث من أشكال الاتصال، عن طريق تحويلها إلى صوت ونص في الوقت الفعلي. كما تعمل ميزة الصوت المحيطي على إبقاء المستخدمين على دراية بما يحيط بهم، حيث ترصد المعطيات الصوتية لتحديد الأصوات المحددة حولهم. وتشمل هذه الأصوات أصوات سيارات الطوارئ وأجهزة الإنذار وأجراس الرنين وأصوات الحيوانات وأصوات الذكور والإناث وأصوات الأطفال وحتى بعض أنواع الموسيقى. وبمجرد اكتشاف هذه الأصوات، يتم تنبيه المستخدمين عن طريق الاهتزاز والإشارات المرئية. وتتيح ميزة الاستماع من جهتها للمستخدمين الصم "قراءة" ما يقوله الآخرون.

يستطيع SpeakLiz فهم ما يصل إلى 48 لغة (في نظام أندرويد) و 35 (في نظام iOS). وعندما يتحدث شخص إلى مستخدم SpeakLiz، يتم تحويل صوته إلى تنسيق النص في الوقت الفعلي. ويمكن للمستخدم قراءة ما يقال والرد وفقاً لذلك. وبهذا يكون المستخدمين الصم الآن قادرين على أداء الأنشطة اليومية دون أي عائق. ونحن نعلم أن التحدث هو عملية معاكسة للاستماع. ومن هنا وباستخدام هذه الميزة، يمكن للمستخدم الصم كتابة أفكارهم. كما يمكن بعد ذلك تحويل النص إلى كلام بضغط زر (بما يشمل الرموز الانفعالية).



الشكل 3: SpeakLiz من تأليف

لقد دخل مركز مدى في شراكة مع سيدستارز (Seedstars) للترويج لحلول النفاذ إلى تكنولوجيا المعلومات والاتصالات بهدف نهائي يتمثل في دعم الشركات الناشئة في مجال النفاذ في قطر وخارجها لتحسين حياة

الأشخاص ذوي الإعاقة. ويتشارك الطرفان الرؤية لدعم الابتكار من خلال العمل مع أفضل الشركات الناشئة في هذا المجال وتقديم التمويل المناسب والخبرة المتخصصة [10].
في عام 2020، فازت شركة تالوف بجائزة النفاذ إلى تكنولوجيا المعلومات والاتصالات الرقمية من مدى- سيدستارز. وفي الأونة الأخيرة، أتاحت تالوف الوصول المجاني إلى تكنولوجيا SpeakLiz و Vision للأشخاص ذوي الإعاقات البصرية والسمعية في الإكوادور. كما فاز الفريق بالجائزة الأولى في أكبر مسابقة لريادة الأعمال الإكوادورية، وتم اختيار الشركة كواحدة من أفضل 200 شركة ناشئة في العالم في مؤتمر قمة الويب بفضل تطبيق SpeakLiz. وفازت الشركة أيضاً بالجائزة الثانية في مسابقة قناة التاريخ (History Channel) كفكرة لتغيير التاريخ بفضل تطبيق Vision.

وقد أثبتت شركة تالوف أن العديد من هذه الحلول التكنولوجية، بما في ذلك التعرف على الإيماءات، قد نجحت في مجال تيسير النفاذ، هذا المجال الذي يشكل التعامل معه واحداً من أكثر السيناريوهات تعقيداً لاستدلال خوارزميات الذكاء الاصطناعي نظراً للعدد الكبير من العوامل التي يجب أخذها في الاعتبار. وبالتالي، فقد حان الوقت للاستفادة من هذه الأداة لتطبيقها بما يتجاوز إمكانية النفاذ. حيث يمكن أن تستفيد العديد من الصناعات الأخرى من أداة كهذه. وفي الواقع، هناك الكثير من الصناعات الأخرى التي لا تدرک حتى كيف يمكن للتعرف على الإيماءات والتقنيات الأخرى أن تؤثر بشكل إيجابي على عملياتها.

الخاتمة

غالباً ما يكون الأسلوب الشائع في هذا المجال للتعرف على لغة الإشارة هو استخدام المثيلات أو الخوادم البعيدة / عبر الإنترنت التي تتمتع بقوة معالجة كبيرة. ومع ذلك، فإن التحدي يكمن في أنه حتى مع وجود اتصال إنترنت فائق السرعة، فإن هناك دائماً فترة كمون. وبالنسبة للمستخدم الحقيقي، يمكن أن يشكل هذا العامل الفرق بين حدوث مشكلة ما أو عدم حدوثها. ويقوم فريق تالوف بإجراء مزيد من البحث لمطابقة الأداء الحالي في الوقت الفعلي مع إمكانية التعلم من أنماط جديدة من تجربة كل مستخدم لتضاف إلى النموذج المركزي. ومن المحتمل أن يتطلب الأمر استخدام بعض المثيلات السحابية لتحديثات نموذج محددة عند الطلب ثم نشر جميع أنماط الشبكات العصبية التلافيفية الجديدة مرة أخرى في بيئة محلية. وبهذا يمكنه الاستمرار في تقديم أداء في الوقت الفعلي باستخدام الذكاء الاصطناعي بشكل مستقل.

المراجع

- [1] S. Ameur, A. B. Khalifa, and M. S. Bouhleb, "A novel hybrid bidirectional unidirectional LSTM network for dynamic hand gesture recognition with leap motion," *Entertain. Comput.*, vol. 35, p. 100373, 2020.
- [2] A. Othman and M. Jemni, "Designing high accuracy statistical machine translation for sign language using parallel corpus: case study English and American Sign Language," *J. Inf. Technol. Res. JITR*, vol. 12, no. 2, pp. 134–158, 2019.
- [3] A. Lahiri, A. Othman, D. A. Al-Thani, and A. Al-Tamimi, "Mada Accessibility and Assistive Technology Glossary: A Digital Resource of Specialized Terms," in *ICCHP*, 2020, p. 207.

- [4] M. Jemni, K. Khribi, A. Othman, O. Elghoul, and K. Jaballah, "AlecsoApps: Toward empowering mobile applications development in the Arab world," in *State-of-the-art and future directions of smart learning*, Springer, 2016, pp. 87–93.
- [5] M. J. Garbade, "Clearing the confusion: Ai vs machine learning vs deep learning differences," *Data Sci.*, vol. 14, 2018.
- [6] H. Cooper, B. Holt, and R. Bowden, "Sign language recognition," in *Visual analysis of humans*, Springer, 2011, pp. 539–562.
- [7] A. Amir *et al.*, "A low power, fully event-based gesture recognition system," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7243–7252.
- [8] L. Zhang, G. Zhu, L. Mei, P. Shen, S. A. A. Shah, and M. Bennamoun, "Attention in convolutional LSTM for gesture recognition," *Adv. Neural Inf. Process. Syst.*, vol. 31, 2018.
- [9] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, "Sign language recognition using convolutional neural networks," in *European Conference on Computer Vision*, 2014, pp. 572–578.
- [10] D. Al Thani, A. Al Tamimi, A. Othman, A. Habib, A. Lahiri, and S. Ahmed, "Mada Innovation Program: A Go-to-Market ecosystem for Arabic Accessibility Solutions," in *2019 7th International conference on ICT & Accessibility (ICTA)*, 2019, pp. 1–3.