الشبكات العصبونية الالتفافية (Convolutional Neural Networks CNNs) هي نوع من الشبكات العصبية تم تصميمها خصيصاً لمهام التعرف على الصور والفيديو. تم تقديمها لأول مرة في الثمانينيات من القرن الماضي ، لكنها لم تنتشر حقًا حتى أوائل عقد 2010، وذلك بفضل توفر مجموعات بيانات كبيرة ، مثل ImageNet ، وتطوير وحدات المعالجة الرسومية القوية التي يمكنها تدريب الشبكات الكبيرة بسرعة.

الفكرة الرئيسية وراء CNNs هي تطبيق مرشحات (المعروفة أيضًا باسم الأوزان أو النوى) على الصورة لاستخراج الميزات في مواقع مختلفة. يتم دمج هذه الميزات في طبقات متعاقبة لتشكيل ميزات عالية المستوى ، وهكذا ، حتى تنتج طبقة الإخراج النهائية مجموعة من التنبؤات لصورة الإدخال.

تتألف معمارية الشبكة العصبية الالتفافية التقليدية من ثلاثة أنواع من الطبقات: طبقات الالتفاف (Convolutional Layers)، وطبقات الاقتراع (Pooling Layers)، وطبقات متصلة بالكامل (Fully Connected Layers). وتقوم طبقات الالتفاف بتطبيق مرشحات على الصورة المدخلة، في حين تقوم طبقات الاقتراع بتخفيض حجم خرائط السمات التي تنتجها طبقات الالتفاف، وهذا يقلل من التكلفة الحسابية للطبقات التالية. وتستخدم طبقات متصلة بالكامل لإنتاج الناتج النهائي، وعادة ما تأتي متبوعة بتابع تفعيل مثل SoftMax الذي يحول الناتج إلى توزيع احتمالي على الفئات (Probability Distribution).

أحد مزايا الشبكات العصبية الالتفافية (CNN) هو قدرتها على تعلم تمثيلات الميزات تلقائياً من البيانات الخام، دون الحاجة إلى تصميم ميزات باليد (Hand-Engineered Features). وهذا يجعلها قادرة على التكيف مع مجموعة واسعة من مهام التعرف على الصور، بدءًا من كشف الكائنات (Object Detection) والشرائح (Segments) إلى وصف الصور (Image Description) ونقل الأسلوب العصبي (Neural Style Transfer).

في السنوات الأخيرة، تم استخدام الشبكات العصبيةالالتفافية (CNNs) لتحقيق أداء يصل إلى أحدث الإنجازات على مجموعة متنوعة من معايير التعرف على الصور، بما في ذلك ImageNet و COCO. وقد تم تطبيقها أيضًا على مجالات أخرى، مثل معالجة اللغة الطبيعية والتعرف على الكلام، حيث حققت نتائج مشجعة.

على الرغم من نجاحها، إلا أنَّ الشبكات العصبونية الالتفافية (CNNs) ليست بلا عيوب. فهي تتطلب كميات كبيرة من بيانات التدريب ويمكن أن تكون حساسة للتغييرات الصغيرة في الإدخال مثل الدوران أو التحجيم. وهي عادة ما تكون باهظة التكلفة حسابياً، خاصة عندما تُطبَّق على الصور أو مقاطع الفيديو ذات الدقة العالية.

في الأقسام التالية، سنستكشف بعض المكونات الرئيسية لشبكات CNN بالتفصيل، بما في ذلك معايير اختيار الشبكة الصحيحة لحل مشكلة البحث المعينة، والمدخلات والطبقات المستخدمة في شبكات CNNs، وتوابع التفعيل Sigmoid و ReLU، وطبقات الاقتراع Pooling Layers، ووصلات الهروب Skip Connections، وطبقات الانحراف المكافئ، وتدريب الشبكات، وتركيبة الشبكة المشهورة UNet.

**معايير اختيار الشبكة العصبية الالتفافية:**

اختيار الشبكة العصبية الالتفافية المناسبة لمشكلة بحثية محددة في الاستدلال بالصور أمر حاسم لتحقيق أداء عالي واستخدام فعال للموارد. هناك عدة مقاييس يمكن استخدامها لتقييم مدى ملاءمة شبكة CNN لمهمة محددة، بما في ذلك الدقة والسرعة واستخدام الذاكرة.

**الدقة Accuracy:** هي ربما أهم المقاييس لقياس أداء شبكة CNN. فهي تقيس مدى قدرة الشبكة على التنبؤ بالنتيجة الصحيحة لمدخل محدد. عند تقييم دقة CNN، من المهم استخدام مجموعة بيانات وبروتوكول تقييم مناسبين يعكسان خصائص التطبيق المستهدف. على سبيل المثال، إذا كانت المهمة تتضمن كشف الكائنات في الصور الطبيعية، يجب تقييم الشبكة على مجموعة بيانات تحتوي على مجموعة متنوعة من الكائنات في مشاهد واقعية، بمجموعة متنوعة من الأحجام ونسب الجوانب للكائنات.

**السرعة Speed:** هي مقياس آخر مهم، خاصة في التطبيقات في الوقت الحقيقي حيث يكون وقت الاستخراج مهماً. تتأثر سرعة شبكة CNN بعدة عوامل، بما في ذلك حجم الشبكة وعدد المعلمات وتعقيد العمليات والأجهزة المستخدمة للاستخراج. لتقييم سرعة CNN، يقيس الباحثون عادة الوقت الذي يستغرقه المعالج لمعالجة صورة مدخل واحدة على منصة عتاد محددة.

**استخدام الذاكرة Memory Usage:** هو مقياس ثالث يمكن استخدامه لتقييم مدى ملاءمة CNN لمشكلة البحث المحددة. يتم تحديد استخدام الذاكرة بحجم الشبكة وكمية الذاكرة المطلوبة لتخزين التفعيلات الوسيطة أثناء الاستخراج. تميل الشبكات الكبيرة التي تحتوي على العديد من الطبقات والمعلمات إلى تتطلب مزيداً من الذاكرة من الشبكات الأصغر، ويمكن أن تصبح هذه المشكلة عاملاً محددًا على الأجهزة المحدودة من حيث الموارد.

المقاييس الأخرى التي يمكن استخدامها لتقييم أداء شبكات CNN تتضمن كفاءة الطاقة والمتانة أمام الهجمات العدائية والتفسيرية. تعد كفاءة الطاقة مهمة للأجهزة التي تعمل بالبطارية ، بينما تعد المتانة أمام الهجمات العدائية مهمة لتطبيقات الأمان. بينما تعد التفسيرية مهمة لفهم كيفية اتخاذ الشبكة لتوقعاتها وتحديد الانحيازات أو الأخطاء المحتملة.

في الواقع ، غالبًا ما يحتاج الباحثون إلى موازنة هذه المقاييس المختلفة اعتمادًا على المتطلبات الخاصة لمشكلة البحث المحددة. على سبيل المثال ، إذا كانت المهمة هي كشف الأمراض النادرة من الصور الطبية ، فقد تكون الدقة هي المقياس الأكثر أهمية ، بينما قد يكون استخدام الذاكرة أقل قلقًا. ومن ناحية أخرى ، إذا كانت المهمة هي تنفيذ تحليلات فيديو في الوقت الحقيقي على جهاز محمول ، فقد يكون السرعة واستخدام الذاكرة هما أكثر أهمية من الدقة.

بالنسبة لتطبيقنا وبسبب محدودية الموارد الحسابية، فقد قمنا باختيار الشبكة على أساس الاستهلاك الأمثل للموارد الحسابية المتاحة.

في المجمل، يتضمن اختيار الشبكة العصبية الالتفافية المناسبة لمشكلة بحث محددة تقييمًا لعدة مقاييس، بما في ذلك الدقة والسرعة واستخدام الذاكرة. يحتاج الباحثون إلى توازن هذه المقاييس بناءً على متطلبات التطبيق المحدد والموارد المتاحة.

**دخل الشبكات العصبية الالتفافية:**

تتكون الشبكات العصبية الالتفافية عادةً من عدة طبقات، يقوم كل منها بعملية معينة على البيانات المدخلة. المدخلات لـCNN عادةً ما تكون صورة ثنائية الأبعاد، على الرغم من أنها يمكن أن تكون حجمًا ثلاثي الأبعاد (على سبيل المثال، الفيديو). هدف CNN هو تعلم تسلسل ميزات تمثل البيانات المدخلة بطريقة مفيدة للمهمة المستهدفة.

**الصورة المدخلة:**

الطبقة الأولى في CNN هي الطبقة المدخلة، والتي تستقبل الصورة المدخلة الخام. يتم تمثيل الصورة المدخلة عادةً على شكل مصفوفة ثلاثية الأبعاد بأبعاد بالشكل التالي (الارتفاع، العرض، القنوات)، حيث تمثل الارتفاع والعرض أبعاد الصورة، والقنوات تمثل قنوات الألوان (مثل قنوات الألوان الأساسية الثلاث -أحمر، أخضر،أزرق- RGB أو صور التقييس الرمادي Grayscale Images) على سبيل المثال، ستمثل صورة RGB بأبعاد 224 × 224 على شكل مصفوفة من الحجم (224, 224, 3).

قبل إدخال الصورة إلى CNN، غالباً ما يتم معالجتها مسبقًا لتطبيع قيم البكسل وتقليل تأثيرات التباين في الإضاءة. تتضمن الخطوات الشائعة للمعالجة المسبقة طرح قيمة البكسل المتوسطة والقسمة على الانحراف المعياري، أو ضمن أغلب التطبيقات العملية، قسمة قيم جميع البكسلات على أقصى قيمة ممكنة للبكسل (255 في حال استخدام نمط البيانات Unsigned Int 8)

**الطبقة الالتفافية Convolutional Layer:** الطبقة الأكثر أهمية في CNN هي الطبقة الترابطية، والتي تقوم بعملية الترابط على البيانات المدخلة. تشمل عملية الترابط سحب فلتر صغير (يسمى أيضًا النواة أو الوزن) عبر الصورة المدخلة وحساب حاصل الضرب الالتفافي بين الفلتر والشق المدخل المقابل في كل موقع.

مخرجات الطبقة الالتفافية هي مجموعة من خرائط الميزات، والتي تمثل وجود أو عدم وجود معالم معينة في مواقع مختلفة في الصورة المدخلة. يتم تحديد عدد خرائط الميزات بواسطة عدد الفلاتر في الطبقة الترابطية، ويتم إنشاء كل خريطة ميزات بواسطة فلتر مختلف. يتم تعلم كل فلتر بواسطة CNN أثناء التدريب للكشف عن نوع معين من الميزات، مثل الحواف، والزوايا، والبقع.

حجم الفلاتر والخطوة (أي، المسافة بين مواقع الفلاتر المجاورة) هي معاملات فوقية يمكن ضبطها للتحكم في حجم خرائط الميزات الناتجة وحقل الاستجابة للشبكة. حجم الفلتر الأكبر والخطوة الأكبر يؤديان إلى حجم إخراج أصغر وحقل استقبال أكبر، بينما يؤدي حجم الفلتر الأصغر والخطوة الأصغر إلى حجم إخراج أكبر وحقل استجابة أصغر.

لتلخيص، المدخل إلى CNN عادةً ما تكون الصورة ثنائية البعد ممثلة بمصفوفة ثلاثية الأبعاد، والطبقة الأكثر أهمية في CNN هي الطبقة الالتفافية، التي تنفذ عملية الترابط على البيانات المدخلة وتولد مجموعة من خرائط الميزات. حجم الفلاتر والخطوة هي معاملات فوقية يمكن ضبطها للتحكم في حجم خرائط الميزات الناتجة وحقل الاستجابة للشبكة.

**وحدة الانحدار الخطي المصححة (ReLU):** هي واحدة من أشهر توابع التصنيف المستخدمة في الشبكات العصبية الالتفافية CNNs تعتبر وظيفة ReLU بسيطة وفعالة من حيث الحسابات، مما يجعلها مثالية للشبكات العصبية العميقة التي تتطلب عدداً كبيراً من المعلمات.

يتم تعريف وظيفة ReLU على النحو التالي:

حيث هو المدخل إلى الدالة، و هو الإخراج. تعيد وظيفة ReLU قيمة المدخل إذا كانت إيجابية، والصفر في غير ذلك. هذا يعني أن ReLU هي وظيفة عتبية تقوم بتعيين القيم السلبية على الصفر.

تابع التفعيل (Rectified Linear Unit -ReLU-) واحداً من التوابع الأكثر استخداماً في الشبكات العصبية الالتفافية (Convolutional Neural Networks -CNNs). ويعتبر تابع التفعيل ReLU بسيط وفعال حسابياً، مما يجعله مثالياً للشبكات العصبية العميقة التي تتطلب عدداً كبيراً من المعاملات.

تابع التفعيل ReLU لها عدة مزايا على الوظائف الأخرى مثل Sigmoid و Tanh واحدة من أهم المزايا هي أنها فعالة حسابياً، حيث تتطلب فقط عملية مقارنة بسيطة وعملية max، هذا يجعلها أسرع في الحساب من Sigmoid و Tanh التي تتطلب عمليات رياضية أكثر تعقيداً.

ومن المزايا الأخرى لـ ReLU هو أنها يمكن أن تساعد في التخفيف من مشكلة اختفاء المشتق (Vanishing Gradient Problem) التي يمكن أن تحدث في الشبكات العصبية العميقة. يحدث مشكلة اختفاء التدرج عندما تصبح قيم التدرجات صغيرة جداً بمجرد أن تمر من خلال الشبكة، مما يمكن أن يؤدي إلى بطء التعلم أو منعه. ويمكن لـ ReLU أن تساعد في منع القيم التدرجية من الانخفاض بشكل كبير لأنها لا تتشبع، وبالتالي لا تقترب من الحد الأقصى أو الحد الأدنى.

وعلاوة على ذلك، يمكن أن يساعد ReLU على تحسين قلة كثافة الشبكة، حيث يقوم بتحويل عدد كبير من التنشيطات إلى الصفر. وهذا يمكن أن يؤدي إلى استخدام أكثر كفاءة لمعلمات الشبكة، حيث يقلل من التكرار في الشبكة.

على الرغم من مزاياه، لدى ReLU أيضًا بعض القيود. واحدة من أهم القيود هي أنه يمكن أن يعاني من مشكلة (Dying ReLU)، حيث تصبح بعض العصبونات غير فعالة بشكل دائم ولا تساهم في إخراج الشبكة. ويمكن أن يحدث ذلك عندما يكون المدخل إلى العصبون سلبياً بشكل متكرر، حيث يصبح معدل التغيير التدرجي الخاص بالعصب صفراً ويتوقف التعلم. وللتخفيف من هذه المشكلة، تم اقتراح العديد من الإصدارات من ReLU، مثل (Leaky ReLU) و (Parametric ReLU) والتي تضيف منحنى قليلاً أو معلمة يمكن تعلمها إلى الجانب السلبي من الدالة.

باختصار، وحدة الانحراف الخطي المصححة (ReLU) هي تابع تفعيل شائع الاستخدام في الشبكات العصبية الالتفافية (CNNs) تعتبر وحدة ReLU فعالة حسابياً وتساعد على التخفيف من مشكلة اختفاء المشتق (Vanishing Gradient Problem) الذي يمكن أن يحدث في الشبكات العصبية العميقة، ويمكن أن تحسن من قلة كثافة الشبكة. ومع ذلك، يمكن أن تعاني من مشكلة (Dying ReLU) مما أدى إلى تطوير متغيرات مثل (Leaky ReLU) و (Parametric ReLU).

Idea #5: Sigmoid Activation Function

Translate this Article to Arabic:

The Sigmoid activation function is a widely used non-linear function in machine learning and deep learning. It is commonly used in Convolutional Neural Networks (CNNs) for its ability to produce a probabilistic output.

Translate this Article to Arabic:

The sigmoid function is defined as:

f(x) = 1 / (1 + e^(-x))

where x is the input to the function and f(x) is the output. The sigmoid function maps any input value to a value between 0 and 1. This makes it ideal for classification tasks where we want to predict the probability of a certain class.

وظيفة التنشيط سيغمويد هي وظيفة غير خطية مستخدمة على نطاق واسع في تعلم الآلة والتعلم العميق. وغالبًا ما تستخدم في الشبكات العصبية التابعة للتباينات (CNNs) بسبب قدرتها على إنتاج إخراج احتمالي.

تعرف وظيفة السيغمويد على النحو التالي:

f (x) = 1 / (1 + e ^ (- x))

حيث x هو المدخل إلى الدالة و f (x) هو الإخراج. ترتبط وظيفة السيغمويد بأي قيمة مدخلة بقيمة بين 0 و 1. وهذا يجعلها مثالية لمهام التصنيف حيث نريد تنبؤ احتمالية فئة معينة.

Translate this Article to Arabic:

The sigmoid function has several advantages. It is a smooth function that is differentiable everywhere, which makes it suitable for optimization techniques like gradient descent. Additionally, it is bounded between 0 and 1, which helps in producing a probability distribution over the classes.

However, the sigmoid function also has some disadvantages. One of the main disadvantages is that it can suffer from the vanishing gradient problem. This problem occurs when the gradient of the sigmoid function becomes very small as the input value moves away from zero. This can make training deep neural networks with sigmoid activation functions difficult, as the gradients can become too small to allow for meaningful updates to the weights.

تتمتع وظيفة السيغمويد بالعديد من المزايا، فهي دالة سلسة يمكن تفريقها في كل مكان مما يجعلها مناسبة لتقنيات الأمثلة مثل التدرج النزولي. بالإضافة إلى ذلك، تحدد بين 0 و 1 مما يساعد في إنتاج توزيع احتمالات على الفئات.

ومع ذلك، فإن وظيفة السيغمويد لديها أيضًا بعض العيوب. واحدة من العيوب الرئيسية هي أنها يمكن أن تعاني من مشكلة اختفاء التدرج. تحدث هذه المشكلة عندما يصبح التدرج لوظيفة السيغمويد صغير جداً عندما يتحرك القيم المدخلة بعيداً عن الصفر. هذا يمكن أن يجعل تدريب الشبكات العصبية العميقة باستخدام وظائف التنشيط السيغمويدية صعبًا، حيث يمكن أن يصبح التدرج صغيرًا جدًا للسماح بتحديثات ذات مغزى للأوزان.

Translate this Article to Arabic:

Another disadvantage of the sigmoid function is that it is not zero-centered. This can lead to problems during training as it can cause the gradients to be biased towards certain directions, leading to slower convergence or suboptimal solutions.

Despite these limitations, the sigmoid activation function remains a popular choice for certain tasks, such as binary classification problems where we want to predict a probability value between 0 and 1.

In summary, the sigmoid activation function is a non-linear function widely used in machine learning and deep learning, particularly in Convolutional Neural Networks (CNNs) for classification tasks. The sigmoid function produces a probabilistic output that is bounded between 0 and 1. While it has some advantages such as smoothness and probabilistic output, it also has some limitations such as the vanishing gradient problem and not being zero-centered.

عيب آخر لوظيفة السيغمويد هو عدم كونها مركزة على الصفر. هذا يمكن أن يؤدي إلى مشاكل أثناء التدريب حيث يمكن أن يتسبب في تحيز التدرجات نحو اتجاهات معينة، مما يؤدي إلى تباطؤ التقارب أو الحلول غير الأمثل.

على الرغم من هذه القيود، تظل وظيفة التنشيط السيغمويد خيارًا شائعًا لبعض المهام، مثل مشكلات التصنيف الثنائي حيث نريد التنبؤ بقيمة احتمالية بين 0 و 1.

وبشكل ملخص، تعتبر وظيفة التنشيط السيغمويد دالة غير خطية تستخدم على نطاق واسع في تعلم الآلة والتعلم العميق، وخاصة في شبكات التباينات العصبية (CNNs) لمهام التصنيف. تنتج وظيفة السيغمويد إخراجًا احتماليًا محصورًا بين 0 و 1. على الرغم من وجود بعض المزايا مثل السلاسة والإخراج الاحتمالي، فإنها تمتلك بعض القيود مثل مشكلة اختفاء التدرج وعدم التركيز على الصفر.

Idea #6: Pooling Layer

Translate this Article to Arabic:

The pooling layer is an important component of Convolutional Neural Networks (CNNs) used for feature extraction. It is a technique used to reduce the spatial dimensions of the input volume, while retaining the most important features.

Pooling layers operate on the output of the convolutional layer, which is a feature map. The purpose of the pooling layer is to downsample the feature map in order to reduce its spatial size, while also retaining the most important features. This helps to reduce the computation required for processing the feature map in the subsequent layers, while also reducing the risk of overfitting.

الطبقة المجمعة هي مكون مهم في الشبكات العصبية التابعة للتحويل الناقل (CNNs) المستخدمة لاستخراج الميزات. إنها تقنية تستخدم لتقليل الأبعاد المكانية لحجم المدخلات، مع الحفاظ على السمات الأكثر أهمية.

تعمل الطبقات المجمعة على إخراج الطبقة التحويلية، والتي تعد خريطة الميزات. ويتمثل هدف الطبقة المجمعة في التخفيض في حجم خريطة الميزات عن طريق تحجيمها، مع الحفاظ على السمات الأكثر أهمية. وهذا يساعد في تقليل الحسابات اللازمة لمعالجة خريطة الميزات في الطبقات التالية، مع تقليل خطر الاستحواذ الزائد.

Translate this Article to Arabic:

There are several types of pooling layers, including max pooling, average pooling, and L2-norm pooling. Max pooling is the most commonly used type of pooling layer. It works by dividing the input feature map into non-overlapping rectangular regions and outputting the maximum value in each region. The size of the regions is determined by the pooling window, which is typically a small square or rectangular region. The output of the max pooling layer is a downsampled feature map that is smaller in spatial size than the input feature map.

Average pooling, on the other hand, works by taking the average value of the pixels in each pooling window. This type of pooling layer can be useful for reducing the impact of outliers in the data. L2-norm pooling is a variation of the average pooling layer that uses the L2-norm instead of the average to compute the output value for each pooling window.

هناك العديد من أنواع الطبقات المجمعة، بما في ذلك التجميع الأقصى، وتجميع المتوسط، وتجميع L2-norm. يعد التجميع الأقصى هو النوع الأكثر استخدامًا من الطبقات المجمعة. يعمل عن طريق تقسيم خريطة الميزات المدخلة إلى مناطق مستطيلة غير متداخلة وإخراج القيمة القصوى في كل منطقة. يتم تحديد حجم المناطق بواسطة نافذة التجميع، التي تكون عادةً منطقة مربعة أو مستطيلة صغيرة. يكون إخراج الطبقة المجمعة الأقصى هو خريطة الميزات المخفضة التي تكون أصغر بحجم مكاني من خريطة الميزات المدخلة.

من ناحية أخرى، يعمل تجميع المتوسط عن طريق أخذ القيمة المتوسطة للبكسلات في كل نافذة التجميع. يمكن أن يكون هذا النوع من الطبقات المجمعة مفيدًا لتقليل تأثير القيم المتطرفة في البيانات. تجميع L2-norm هو تباين لطبقة التجميع المتوسطة يستخدم L2-norm بدلاً من المتوسط لحساب القيمة الناتجة لكل نافذة تجميع.

Translate this Article to Arabic:

The pooling layer plays a critical role in the overall performance of the CNN. It helps to reduce the spatial dimensions of the feature map, which reduces the number of parameters that need to be learned by the network. This, in turn, reduces the risk of overfitting and makes the network more computationally efficient.

One potential disadvantage of pooling layers is that they can result in the loss of some spatial information. Since pooling involves downsampling the feature map, some of the spatial details can be lost in the process. This can be mitigated by using larger strides in the convolutional layers or by using other techniques such as dilated convolutions.

تلعب طبقة التجميع دورًا حاسمًا في الأداء العام لشبكات العصب العصبي التكراري. إنها تساعد على تقليل الأبعاد المكانية لخريطة الميزات، مما يقلل من عدد المعلمات التي يجب أن يتعلمها الشبكة. وبالتالي، يتم تقليل مخاطر الانحراف الزائد وجعل الشبكة أكثر كفاءة حسابية.

واحدة من المساوئ المحتملة لطبقات التجميع هي أنها يمكن أن تؤدي إلى فقدان بعض المعلومات المكانية. نظرًا لأن التجميع ينطوي على تخفيض أبعاد خريطة الميزات، يمكن فقدان بعض التفاصيل المكانية في العملية. يمكن التخفيف من ذلك عن طريق استخدام خطوات أكبر في الطبقات الترددية أو باستخدام تقنيات أخرى مثل الترددات المتباعدة.

Translate this Article to Arabic:

In summary, the pooling layer is a critical component of Convolutional Neural Networks used for feature extraction. It helps to reduce the spatial dimensions of the input feature map, while retaining the most important features. There are several types of pooling layers, including max pooling, average pooling, and L2-norm pooling. The pooling layer plays a critical role in the overall performance of the CNN, but care must be taken to mitigate the loss of spatial information that can occur during pooling.

لخلاصة الفقرة، طبقة التجميع هي عنصر حاسم في شبكات العصب العصبي التكراري المستخدمة لاستخراج الميزات. إنها تساعد على تقليل الأبعاد المكانية لخريطة الميزات الإدخالية، مع الاحتفاظ بالميزات الأكثر أهمية. هناك عدة أنواع من طبقات التجميع، بما في ذلك التجميع الأقصى والتجميع المتوسط ​​والتجميع L2-norm. تلعب طبقة التجميع دورًا حاسمًا في الأداء العام لشبكات العصب العصبي التكراري، ولكن يجب اتخاذ الحيطة والحذر للحد من فقدان المعلومات المكانية التي يمكن أن تحدث خلال عملية التجميع.

Idea #7: Skip Connections

Translate this Article to Arabic:

Skip connections, also known as residual connections, are a type of shortcut connection used in deep neural networks. They were first introduced in the ResNet architecture, which achieved state-of-the-art performance on several computer vision tasks, including image classification and object detection.

In a neural network, skip connections allow information to be passed from one layer to another without going through the intermediate layers. This can be useful for several reasons. First, it can help to alleviate the problem of vanishing gradients, which can occur in deep networks when the gradients become too small to effectively update the weights. By allowing information to bypass the intermediate layers, skip connections can help to maintain the gradients and improve the overall performance of the network.

تعتبر الاتصالات المتخطية، المعروفة أيضاً باسم الاتصالات الباقية، نوعًا من الاتصالات القصيرة المستخدمة في الشبكات العصبية العميقة. وقد تم تقديمها لأول مرة في بنية ResNet، التي حققت أداءًا متفوقًا في العديد من المهام الرئيسية لرؤية الحاسوب، بما في ذلك تصنيف الصور والكشف عن الكائنات.

في الشبكة العصبية، تسمح الاتصالات المتخطية للمعلومات بالانتقال من طبقة إلى أخرى دون الذهاب من خلال الطبقات الوسيطة. ويمكن أن يكون هذا مفيدًا لعدة أسباب. أولاً، يمكن أن يساعد في التخفيف من مشكلة اختفاء التدرجات، والتي يمكن أن تحدث في الشبكات العميقة عندما تصبح التدرجات صغيرة جدًا لتحديث الأوزان بشكل فعال. من خلال السماح للمعلومات بالتجاوز من الطبقات الوسيطة، يمكن للاتصالات المتخطية المساعدة في الحفاظ على التدرجات وتحسين الأداء العام للشبكة.

Translate this Article to Arabic:

Second, skip connections can also help to improve the accuracy of the network by providing more direct connections between the input and output. This can be especially useful for tasks where the input and output have a strong correlation, such as image segmentation or image-to-image translation.

There are several types of skip connections, including identity skip connections, projection skip connections, and dense skip connections. Identity skip connections are the simplest type, and involve adding the input to the output of a convolutional layer. This can help to maintain the spatial resolution of the feature map and improve the performance of the network.

بالإضافة إلى ذلك، يمكن أن تساعد الاتصالات المتخطية على تحسين دقة الشبكة عن طريق توفير اتصالات أكثر مباشرة بين المدخلات والمخرجات. ويمكن أن يكون هذا مفيدًا بشكل خاص للمهام التي تتمتع المدخلات والمخرجات فيها بترابط قوي، مثل تجزئة الصور أو ترجمة الصورة إلى صورة أخرى.

هناك العديد من أنواع الاتصالات المتخطية، بما في ذلك اتصالات المرور الهوية، واتصالات المرور المتوقعة، واتصالات المرور الكثيفة. وتعتبر اتصالات المرور الهوية هي النوع الأبسط، وتنطوي على إضافة المدخلات إلى مخرج طبقة التحويل التكاملي للترددات. ويمكن أن تساعد هذه الطريقة في الحفاظ على الدقة المكانية لخريطة المعالم وتحسين أداء الشبكة.

Translate this Article to Arabic:

Projection skip connections, on the other hand, involve using a convolutional layer to project the input onto the same dimension as the output. This can be useful for cases where the input and output have different spatial resolutions or feature dimensions.

Dense skip connections are a more advanced type of skip connection that involve connecting each layer to every other layer in the network. This can help to improve the information flow between the layers and reduce the risk of information bottlenecks.

أما بالنسبة لاتصالات المرور المتوقعة، فتتضمن استخدام طبقة تحويل تكاملي للترددات لإسقاط المدخلات على نفس الأبعاد كمخرج الشبكة. ويمكن أن تكون هذه الطريقة مفيدة في الحالات التي تختلف فيها الدقة المكانية أو أبعاد المعالم بين المدخلات والمخرجات.

وتعد اتصالات المرور الكثيفة هي نوع متقدم أكثر من اتصالات المرور المتخطية، وتنطوي على ربط كل طبقة بكل طبقة أخرى في الشبكة. ويمكن أن تساعد هذه الطريقة في تحسين تدفق المعلومات بين الطبقات وتقليل مخاطر حدوث الانسداد في معلومات الشبكة.

Translate this Article to Arabic:

Skip connections have become an important component of many state-of-the-art deep neural network architectures, including ResNet, DenseNet, and U-Net. They have been shown to be effective for a wide range of computer vision tasks, including image classification, object detection, semantic segmentation, and image-to-image translation.

In summary, skip connections are a type of shortcut connection used in deep neural networks. They can help to alleviate the problem of vanishing gradients, improve the accuracy of the network, and provide more direct connections between the input and output. There are several types of skip connections, including identity skip connections, projection skip connections, and dense skip connections, and they have become an important component of many state-of-the-art deep neural network architectures.

أصبحت اتصالات المرور المتخطية عنصرًا مهمًا في العديد من بنى الشبكات العصبية العميقة الحديثة، بما في ذلك ResNet و DenseNet و U-Net. وقد ثبت أنها فعالة لمجموعة واسعة من مهام رؤية الحاسوب، بما في ذلك التصنيف الصوري، وكشف الكائنات، والتجزئة الدلالية، والترجمة من صورة إلى صورة.

وبإيجاز، فإن اتصالات المرور المتخطية هي نوع من اتصالات الاختصار التي تستخدم في شبكات العصبية العميقة، ويمكن أن تساعد في تخفيف مشكلة اختفاء التدرجات، وتحسين دقة الشبكة، وتوفير اتصالات مباشرة أكثر بين المدخلات والمخرجات. وتتضمن أنواعًا مختلفة من اتصالات المرور المتخطية، بما في ذلك اتصالات المرور المتخطية المماثلة، واتصالات المرور المتوقعة، واتصالات المرور الكثيفة، وأصبحت عنصرًا مهمًا في العديد من بنى الشبكات العصبية العميقة الحديثة.

Idea #9: Network Training

Translate this Article to Arabic:

Convolutional neural networks (CNNs) are a powerful class of deep learning models used for image recognition, segmentation, and other computer vision tasks. However, training CNNs can be a challenging task, requiring a significant amount of data, computational resources, and hyperparameter tuning.

In this article, we will discuss the basics of CNN training, including the key components of the training process and common techniques used to improve model performance.

Translate this Article to Arabic:

The training process of a CNN typically involves the following steps:

1. Data preparation: The first step in training a CNN is to prepare the data. This may involve collecting and cleaning data, resizing images, and splitting the data into training, validation, and testing sets.

2. Model selection: Once the data is prepared, the next step is to select a CNN architecture that is appropriate for the task at hand. This may involve choosing a pre-trained model, such as VGG, ResNet, or Inception, or designing a custom model from scratch.

3. Loss function: The loss function is a key component of the training process, as it measures the difference between the predicted and true labels for each image in the training set. Common loss functions used in CNNs include categorical cross-entropy for multi-class classification, binary cross-entropy for binary classification, and mean squared error for regression tasks.

4. Optimization algorithm: The optimization algorithm is responsible for updating the model parameters to minimize the loss function during training. Common optimization algorithms used in CNNs include stochastic gradient descent (SGD), Adam, and Adagrad.

5. Hyperparameter tuning: Finally, the performance of the model can be improved by tuning the hyperparameters, such as learning rate, batch size, number of epochs, and regularization strength. This can be done using a grid search or random search approach, or by using more advanced techniques such as Bayesian optimization or genetic algorithms.

عملية تدريب CNN تتضمن عادة الخطوات التالية:

1. إعداد البيانات: الخطوة الأولى في تدريب CNN هي إعداد البيانات. وقد يتطلب هذا جمع البيانات وتنظيفها، وتغيير حجم الصور، وتقسيم البيانات إلى مجموعات التدريب والتحقق والاختبار.
2. اختيار النموذج: بمجرد أن تم إعداد البيانات، فإن الخطوة التالية هي اختيار بنية CNN مناسبة للمهمة المطلوبة. وقد يتضمن ذلك اختيار نموذج مدرب مسبقًا مثل VGG أو ResNet أو Inception، أو تصميم نموذج مخصص من البداية.
3. وظيفة الخسارة: تعد وظيفة الخسارة مكونًا رئيسيًا في عملية التدريب، حيث تقيس الفرق بين التصنيفات المتوقعة والفعلية لكل صورة في مجموعة التدريب. وتشمل الوظائف الشائعة المستخدمة في CNNs خسارة التبويب الفئوية للتصنيف المتعدد، وخسارة التبويب الثنائية للتصنيف الثنائي، وخطأ المربعات المتوسطة لمهام الانحدار.
4. خوارزمية التحسين: تكون خوارزمية التحسين مسؤولة عن تحديث معلمات النموذج لتقليل وظيفة الخسارة خلال التدريب. وتشمل الخوارزميات الشائعة المستخدمة في CNNs الانحدار التدرجي العشوائي (SGD) و Adam و Adagrad.
5. ضبط الهايبرباراميتر: يمكن تحسين أداء النموذج في النهاية من خلال ضبط الهايبرباراميترات، مثل معدل التعلم وحجم الدفعة وعدد الدورات التدريبية وقوة التنظيم. يمكن القيام بذلك باستخدام البحث الشبكي Grid Search أو البحث العشوائي، أو باستخدام خوارزميات أكثر تقدماً مثل التحسين البايزي Bayesian Optimization أو الخوارزميات الجينية.

Translate this Article to Arabic:

To monitor the progress of the training process, several metrics can be used, including the loss function, accuracy, precision, recall, and F1 score. These metrics can be computed on the training, validation, and testing sets to evaluate the performance of the model on different data.

In addition to the basic components of CNN training, several techniques have been proposed to improve model performance, including:

Translate this Article to Arabic:

1. Data augmentation: Data augmentation involves generating new training data by applying random transformations to the original images, such as rotations, translations, and flips. This can help prevent overfitting and improve model generalization.

2. Transfer learning: Transfer learning involves using a pre-trained CNN model on a large dataset, such as ImageNet, as a starting point for a new task. The pre-trained model can be fine-tuned on the new task, or used as a fixed feature extractor.

Translate this Article to Arabic:

3. Regularization: Regularization techniques, such as L1 or L2 regularization, dropout, or batch normalization, can be used to prevent overfitting and improve model generalization.

4. Early stopping: Early stopping involves monitoring the validation loss during training and stopping the training process when the loss starts to increase, indicating that the model is overfitting.

1. تكبير البيانات: يتضمن تكبير البيانات إنشاء بيانات تدريب جديدة عن طريق تطبيق تحويلات عشوائية على الصور الأصلية ، مثل الدوران والترجمة والاعتدال. يمكن أن يساعد هذا في منع الاحتكاك الزائد وتحسين تعميم النموذج.

2. التعلم النقلي: يتضمن التعلم النقلي استخدام نموذج CNN مدرب مسبقًا على مجموعة بيانات كبيرة ، مثل ImageNet ، كنقطة انطلاق لمهمة جديدة. يمكن ضبط النموذج المدرب مسبقًا على المهمة الجديدة أو استخدامه كمستخرج للميزات الثابت.

3. التنظيم: يمكن استخدام تقنيات التنظيم ، مثل التنظيم L1 أو L2 أو قطر أو تطبيع الدفعات ، لمنع الاحتكاك الزائد وتحسين تعميم النموذج.

4. التوقف المبكر: يتضمن التوقف المبكر مراقبة خسارة التحقق خلال التدريب وإيقاف عملية التدريب عندما تبدأ الخسارة في الارتفاع ، مما يشير إلى أن النموذج يعاني من الاحتكاك الزائد.

Translate this Article to Arabic:

In conclusion, training a CNN involves several key components, including data preparation, model selection, loss function, optimization algorithm, and hyperparameter tuning. By using these components in combination with advanced techniques such as data augmentation, transfer learning, and regularization, it is possible to achieve state-of-the-art performance on a wide range of computer vision tasks.

في الخلاصة، يتضمن تدريب شبكات التعلم العميق للتعرف على الصور (CNN) عدة مكونات رئيسية، بما في ذلك إعداد البيانات، اختيار النموذج، وظيفة الخسارة، خوارزمية الأمثلية، وضبط الهايبربارامترات. باستخدام هذه المكونات بالاشتراك مع التقنيات المتقدمة مثل تعزيز البيانات، والتعلم النقلي، والتعويض، يمكن تحقيق أداء رائع في مجموعة واسعة من مهام رؤية الحاسوب.

Idea #10: Network Architecture of the Famous UNet

Translate this Article to Arabic:

In recent years, deep learning models have been used extensively in the field of medical image segmentation. Among these models, the U-Net has become one of the most popular models, and it has been widely used for various medical imaging applications.

The U-Net is a convolutional neural network architecture that was introduced by Ronneberger et al. in 2015 for biomedical image segmentation. It was specifically designed for medical image analysis tasks, such as segmenting cells, tissues, and organs in images from microscopy and other medical imaging modalities. The name "U-Net" comes from the architecture's shape, which is symmetric and U-shaped.

في السنوات الأخيرة ، تم استخدام نماذج التعلم العميق على نطاق واسع في مجال تجزئة الصور الطبية. ومن بين هذه النماذج ، أصبح U-Net واحدًا من أشهر النماذج ، وتم استخدامه على نطاق واسع لتطبيقات التصوير الطبي المختلفة.

U-Net هو بنية للشبكات العصبية التحويلية تم تقديمها من قبل Ronneberger وآخرين في عام 2015 لتجزئة الصور الحيوية. تم تصميمه خصيصًا لمهام تحليل الصور الطبية ، مثل تجزئة الخلايا والأنسجة والأعضاء في الصور من المجهريات وغيرها من أنواع صور التصوير الطبي. يأتي اسم "U-Net" من شكل البنية الذي هو متماثل وعلى شكل حرف U.

Translate this Article to Arabic:

The U-Net architecture is composed of two main parts: the contracting path and the expansive path. The contracting path is used to capture the context of the input image by applying multiple convolution and pooling operations, which reduce the spatial resolution of the feature maps while increasing their number. The expansive path, on the other hand, is used to recover the original spatial resolution of the input image by applying multiple upsampling and convolution operations.

The contracting path consists of several layers, each of which performs two consecutive 3x3 convolutions followed by a rectified linear unit (ReLU) activation function, and then a 2x2 max-pooling operation. The number of feature maps in each layer is gradually increased while the spatial resolution is reduced. This process results in a high-level feature representation of the input image.

تتكون بنية U-Net من جزئين رئيسيين: المسار الانقباضي والمسار التمددي. يستخدم المسار الانقباضي لالتقاط سياق الصورة المدخلة عن طريق تطبيق عدة عمليات تحويل وتجميع، والتي تقلل من الدقة المكانية لخرائط المعالم مع زيادة عددها. يستخدم المسار التمددي ، بالمقابل ، لاستعادة الدقة المكانية الأصلية للصورة المدخلة عن طريق تطبيق عدة عمليات رفع الدقة والتحويل.

يتكون المسار الانقباضي من عدة طبقات، حيث ينفذ كل منها عمليتي تحويل بمقاس 3x3 متتاليتين، تليها وظيفة تفعيل خطية (ReLU) ثم عملية تجميع بأقصى قيمة بمقاس 2x2. يتم زيادة عدد خرائط المعالم في كل طبقة تدريجيًا في حين يتم تقليل الدقة المكانية. يؤدي هذا العملية إلى تمثيل معالم عالية المستوى للصورة المدخلة.

Translate this Article to Arabic:

The expansive path consists of several layers, each of which performs an up-convolution operation followed by a concatenation with the corresponding feature map from the contracting path. Then, two consecutive 3x3 convolutions are applied, followed by a ReLU activation function. This process allows the network to recover the spatial resolution of the input image while preserving the semantic information learned in the contracting path.

In addition to the contracting and expansive paths, the U-Net also includes skip connections between the corresponding layers of the contracting and expansive paths. These skip connections allow the network to use the high-level features learned in the contracting path to refine the segmentation in the expansive path.

يتكون المسار التمددي من عدة طبقات، حيث ينفذ كل منها عملية تحويل رفع الدقة متبوعة بالاتحاد مع خريطة المعالم المقابلة من المسار الانقباضي. ثم يتم تطبيق عمليتي تحويل بمقاس 3x3 متتاليتين، تليها وظيفة تفعيل خطية (ReLU). يسمح هذا العملية بإعادة استعادة الدقة المكانية للصورة المدخلة مع الحفاظ على المعلومات الدلالية المستخرجة في المسار الانقباضي.

بالإضافة إلى المسار الانقباضي والمسار التمددي، يتضمن U-Net أيضًا اتصالات تجاوز بين الطبقات المقابلة في المسار الانقباضي والمسار التمددي. تسمح هذه الاتصالات بإستخدام المعالم عالية المستوى المستخرجة في المسار الانقباضي لتحسين التفصيل في المسار التمددي.

Translate this Article to Arabic:

The U-Net has become popular in the medical imaging community due to its outstanding performance in various segmentation tasks. It has been shown to achieve state-of-the-art results on several medical image datasets, including the ISBI cell tracking challenge, the EM segmentation challenge, and the MICCAI 2015 prostate segmentation challenge.

To sum up, the U-Net is a powerful convolutional neural network architecture that has become the go-to choice for many medical image segmentation tasks. Its symmetric and U-shaped architecture, as well as its contracting and expansive paths, allow it to capture the context of the input image while preserving the spatial resolution of the output segmentation. With the help of skip connections, the U-Net is able to refine the segmentation by leveraging the high-level features learned in the contracting path.

أصبح U-Net شائعًا في مجتمع التصوير الطبي بسبب أدائه المتميز في مهام التجزئة المختلفة. وقد أظهرت النتائج الأولية أن U-Net تحقق نتائج عالية الجودة على عدة مجموعات بيانات لصور طبية، بما في ذلك تحدِّي تتبع الخلايا ISBI وتحدِّي تجزئة EM وتحدِّي تجزئة البروستاتا MICCAI 2015.

وبشكل عام، فإن U-Net هو تصميم شبكة عصبونية تحويلية قوية أصبحت خيارًا رائجًا للعديد من مهام تجزئة الصور الطبية. وتمكن هذا التصميم من التقاط سياق الصورة المدخلة مع الحفاظ على الدقة المكانية لتجزئة الإخراج بفضل التصميم الرمزي المتناظر والمستلهم من الحرف الألماني "يو". كما يتيح للشبكة الانقباضية والتمددية التقاط المعلومات الدلالية المهمة للتفاصيل واستعادة الدقة المكانية بالترتيب. ويستخدم نظام الاتصالات التجاوزية لتحسين تجزئة الصورة من خلال الاستفادة من المعلومات الدلالية العالية المستوى التي تم استخراجها في المسار الانقباضي.