

# شناسایی ترک در سازه‌های مصالح بنایی به کمک بینایی رایانه براساس یادگیری عمیق

محمد موسوی (دانشجوی کارشناسی ارشد)

علی بخشی\* (دانشیار)

دانشکده‌ی مهندسی عمران، دانشگاه صنعتی شریف

مهندسی عمران شریف، تابستان ۱۴۰۱ (۳۸-۲، شماره ۲/۱، ص. ۱۰۸-۹۹، پژوهشی)

سازه‌های مصالح بنایی، شمار بالایی از سازه‌های ساخته‌ی دست بشر را تشکیل می‌دهند. با توجه به اهمیت تاریخی و همچنین شرایط محیطی، بسیاری از سازه‌های مذکور نیازمند یک سیستم پایش سلامت کارا هستند. پیشرفت‌های اخیر در حوزه‌ی هوش مصنوعی و بینایی رایانه، دستیابی به سیستم موردنظر را بیش از پیش نزدیک کرده است. در پژوهش حاضر، فرایند شناسایی ترک در تصاویر سازه‌های مصالح بنایی با استفاده از یک مدل یادگیری عمیق انجام پذیرفته است. در توسعه‌ی الگوریتم مذکور، از یک مدل رمزگذار-رمزگشای کانولوشنال استفاده شده است که خروجی آن، تصویری بخش‌بندی شده از تصویر ورودی بوده است. در پژوهش حاضر، مجموعه‌ی داده‌ی متشکل از ۱۱۵ تصویر از سازه‌های مصالح بنایی که وقوع ترک در آن‌ها رخ داده است، گردآوری و پس از برچسب‌گذاری دستی آن‌ها، اقدام به آموزش و ارزیابی شده است. در نهایت، دقت‌های صحت، فراخوانی و معیار  $F1$  به ترتیب برابر  $0.81/0.44$ ،  $0.71/0.41$  و  $0.75/0.36$  برای داده‌های ارزیابی و در مقیاس پیکسل حاصل شده است.

واژگان کلیدی: پایش سلامت سازه‌ها، یادگیری عمیق، بینایی رایانه، شبکه‌های عصبی کانولوشن، بخش‌بندی تصویر.

mohammad.mousavi@sharif.edu  
bakhshi@sharif.edu

## ۱. مقدمه

مصالح بنایی از مصالح بسیار پرکاربرد در مباحث ساختمان از گذشته تا به امروز است. در سازه‌های با قدمت بیشتر، از مصالح بنایی به عنوان عضو سازه‌ی نیز استفاده شده است که عمومیت ساختمان‌ها و سازه‌های مصالح بنایی را شامل می‌شوند. اگرچه به دلیل آسیب‌پذیری بسیار بالای سازه‌های مصالح بنایی در مقابل زلزله، استفاده از آن در مناطق زلزله‌خیز کاهش پیدا کرده است؛ همچنان استفاده از مصالح بنایی و آجر در بسیاری از مناطق با لرزه‌خیزی پایین رواج دارد. بسیاری از ساختمان‌های با قدمت زیاد وجود دارند که آجر عنصر کلیدی مصالح آن‌هاست، ولی گذشت زمان و بارگذاری‌های خارج از انتظار باعث ضعف عملکرد آن‌ها شده است. در نتیجه، در مناطق زلزله‌خیز چون ایران، برای حصول اطمینان از سلامت سازه‌های مصالح بنایی در گذر زمان و همچنین در مواجهه با بارگذاری‌های شدید چون زلزله، نیاز به بازرسی و پایش مداوم ساختمان‌های با قدمت زیاد محسوس است. تاکنون مؤثرترین روش موجود برای پایش‌های ذکر شده، بازرسی انسانی بوده است؛ اما خطاهای موجود و هزینه‌ی گزاف روش اخیر، همواره آن را با چالش همراه

با توجه به موارد ذکر شده‌ی اخیر، نیاز به روشی جایگزین برای پایش سازه‌های با قدمت زیاد بسیار محسوس است. بینایی رایانه یکی از روش‌های محبوب در مطالعات است که علاوه بر اینکه از هزینه‌ها و مخاطرات بازرسی انسانی مصون است، از روش‌های غیرمخرب در ارزیابی سازه‌هاست، که برای سازه‌های مصالح بنایی که اغلب سازه‌های با قدمت زیاد هستند، بسیار مطلوب و مناسب است.

اخیراً پیشرفت‌های گسترده‌ی هوش مصنوعی در حوزه‌ی یادگیری عمیق<sup>۱</sup>، مرزهای جدیدی در مبحث بینایی رایانه ایجاد کرده است. اصلی‌ترین ابزاری که یادگیری عمیق در اختیار بینایی رایانه قرار داده است، شبکه‌های عصبی کانولوشن<sup>۲</sup> هستند. کریشنسکی<sup>۳</sup> و همکاران (۲۰۱۲)،<sup>۱</sup> با استفاده از یک شبکه‌ی عصبی کانولوشنال عمیق توانستند به دقت  $0.84/0.7$  در دسته‌بندی<sup>۴</sup> تصاویر رقابت ایمیجنت<sup>۵</sup> که اصلی‌ترین رقابت در حوزه‌ی بینایی رایانه است، دست پیدا کنند. این موضوع نقطه‌ی

\* نویسنده مسئول

تاریخ: دریافت ۱۴۰۰/۹/۲۷، اصلاحیه ۱۴۰۰/۱۱/۳۰، پذیرش ۱۴۰۰/۱۲/۲۱.

DOI:10.24200/J30.2022.59496.3055

آغازی جدید در مبحث بینایی رایانه بوده و پس از آن تقریباً تمامی مسائل حوزه‌ی مذکور با کمک یادگیری عمیق قابلیت حل داشته‌اند. چاه<sup>۶</sup> و همکاران (۲۰۱۷)،<sup>[۳]</sup> از یک شبکه‌ی عصبی کانولوشن برای دسته‌بندی تصاویر ترک بر روی سطوح بتنی استفاده کردند و دقت ۹۸٪ را برای طبقه‌بندی مذکور گزارش دادند. ایشان پس از آن و در مطالعه‌ی دیگری (۲۰۱۸)،<sup>[۴]</sup> اقدام به شناسایی و دسته‌بندی همزمان پنج نوع آسیب سازه‌ی کردند و به دقت میانگین ۸۷/۷٪ دست یافتند. محتشم‌خانی و همکاران (۲۰۲۰)،<sup>[۵]</sup> با استفاده از یک مدل کانولوشنال و با بهره‌گیری از یادگیری انتقالی<sup>۷</sup>، اقدام به دسته‌بندی تصاویر ترک برای داده‌های مربوط توربین‌های گازی کردند و دقت ۹۶/۲۶٪ را گزارش دادند. فنگ<sup>۸</sup> و همکاران<sup>[۶]</sup> (۲۰۱۹)، نیز اقدام به دسته‌بندی تصاویر برای شناسایی آسیب‌های ایجاد شده در سطح سازه‌های آبی کردند که با دقتی بالغ بر ۹۶٪ همراه بود. تمامی مطالعات اخیر از دقت بسیار بالایی شبکه‌های عصبی کانولوشن و یادگیری عمیق در حوزه‌ی دسته‌بندی تصاویر مربوط ترک و آسیب‌های سازه‌ی حکایت دارند. اگرچه دسته‌بندی تصاویر یکی از مسائل بسیار مهم در مبحث بینایی رایانه است، اما در حوزه‌ی پیش سلامت سازه‌ها<sup>۹</sup> اطلاعات کامل و مناسبی در اختیار پژوهشگران قرار نمی‌دهد. با توجه به اهمیت موضوع، در آسیب‌های سازه‌ی، روشی مطلوب نظر است که بتواند علاوه بر تشخیص وجود ترک یا آسیب در تصویر، همزمان جزئیاتی از ابعاد و الگوی آسیب نیز در اختیار پژوهشگران قرار دهد. اگر از منظر بینایی رایانه به مسئله نگاه شود، در مجموع سه نوع مسئله به واسطه‌ی بینایی رایانه حل‌شدنی هستند: (۱) دسته‌بندی تصاویر که معمولاً وجود یا نبود یک شی یا پدیده در تصویر را مشخص می‌کند؛ (۲) تشخیص اشیاء<sup>۱۰</sup> که علاوه بر تشخیص وجود یک شیء تقریبی، محل وجود آن را نیز در اختیار کاربر قرار می‌دهد؛ (۳) بخش‌بندی یا تقسیم‌بندی معنایی<sup>۱۱</sup> تصاویر که یکایک پیکسل‌های موجود در تصویر را دسته‌بندی می‌کند و به هر یک از آن‌ها برچسبی مخصوص اختصاص می‌دهد. بخش‌بندی تصاویر، دقت بسیار بالایی دارد و مسائلی که به واسطه‌ی آن حل می‌شوند، شامل بیشترین جزئیات خواهند بود؛ اما چنین جزئیات و دقتی، هزینه‌ی زیادی نیز با خود به همراه دارد. نخست در بحث جمع‌آوری داده برای آموزش مدل‌های روش مذکور، نیاز است که تمامی داده‌ها به صورت دستی برچسب‌گذاری شوند و لذا فراهم‌سازی هر داده برای آموزش چنین مدلی، ده‌ها برابر روشی چون دسته‌بندی، زمان خواهد برد. اگرچه هزینه‌ی محاسباتی روش نیز می‌تواند به عنوان یکی از مشکلات آن مطرح شود. توسعه‌ی روزافزون واحدهای سخت‌افزاری و همچنین بهینه‌سازی‌های گسترده در شبکه‌های ارائه شده در حوزه‌ی بینایی رایانه تا حد زیادی مشکل اخیر را مرتفع کرده است.

در حوزه‌ی یادگیری عمیق در سال ۲۰۱۵، با ارائه‌ی شبکه‌های عصبی تماماً کانولوشنال<sup>[۷]</sup> مسیر جدیدی در حل مسائل بخش‌بندی تصاویر گشوده شد. یانگ<sup>۱۳</sup> و همکاران (۲۰۱۸)،<sup>[۸]</sup> از شبکه‌های عصبی تماماً کانولوشنال استفاده کردند و فرایند تقسیم‌بندی ترک برای تصویر اعضاء بتنی را به انجام رساندند. البته داده‌ی استفاده شده در پژوهش ایشان که در حقیقت در پژوهشی دیگر و برای عملیات دسته‌بندی ارائه شده بود،<sup>[۹]</sup> پیچیدگی بسیار کمی داشته است، به طوری که با تصاویر واقعی از ترک، اختلاف بسیاری داشته و احتمال افت دقت مدل در مواجهه با مسائل واقعی بسیار زیاد بوده است.

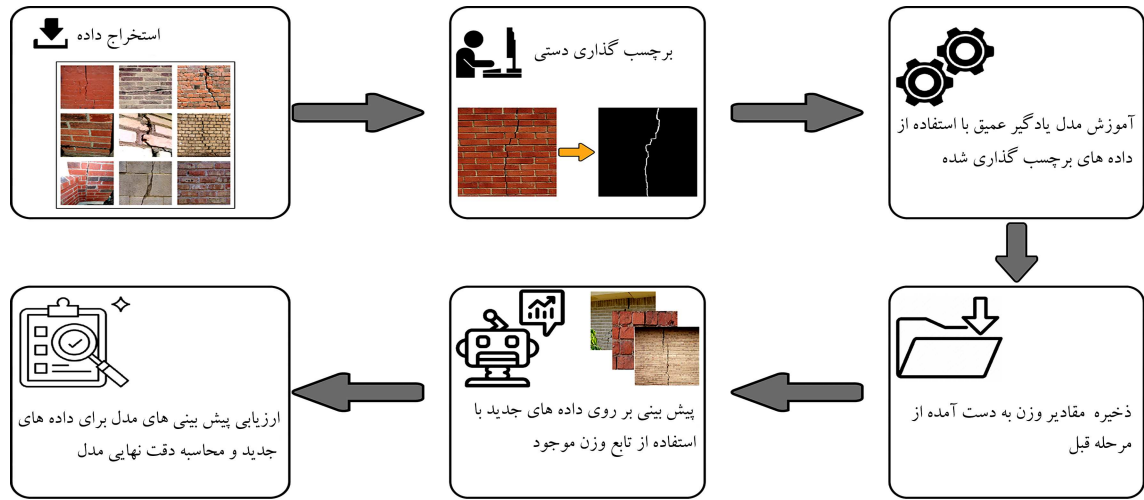
از دیگر مطالعات جدید در حوزه‌ی دسته‌بندی تصاویر می‌توان به U-Net<sup>[۱۰]</sup> اشاره کرد که مشابه شبکه‌های عصبی تماماً کانولوشنال، نوعی الگوریتم رمزگذار-رمزگشا<sup>۱۴</sup> است؛ با این تفاوت که علاوه بر آموزش رمزگذار و رمزگشا، رابطه‌ی بین عناصر آن‌ها نیز به مدل آموزش داده می‌شود. مطالعات بسیاری از مدل مذکور در حل مسائل مهندسی استفاده کرده‌اند که یکی از مسائل مذکور مربوط به شناسایی

سطوح دارای ترک بوده است. لیو<sup>۱۵</sup> و همکاران (۲۰۱۹)،<sup>[۱۱]</sup> با استفاده از مدل اصلی پیشنهادی U-Net<sup>[۱۰]</sup> مدلی ارائه کردند و توانستند با دقتی نزدیک به ۹۰٪، عملیات بخش‌بندی را برای تصاویر ترک بتن به انجام برسانند. ایشان همچنین نشان دادند که مدل مذکور نسبت به مدل‌های پیشین نیاز به داده‌ی آموزش بسیار کمتری دارد (۵۷ داده برای آموزش، در مقایسه با مدل‌های پیشین که دست‌کم از ۴۰۰ داده برای آموزش استفاده کرده بودند).

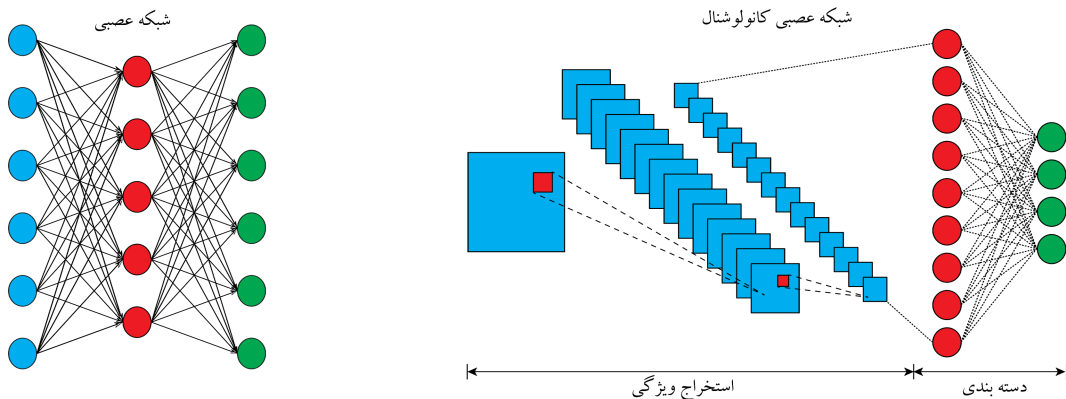
در مطالعه‌ی حاضر قصد بر این بوده است که ضمن جمع‌آوری و برچسب‌گذاری تعداد مناسبی داده‌ی تصویری از ساختمان‌های مصالح بتنی دارای ترک که عمدتاً آجری بودند، نسبت به بخش‌بندی ترک‌ها در تصاویر ذکر شده با استفاده از یادگیری عمیق اقدام شود. مدل استفاده شده در پژوهش حاضر بر پایه‌ی U-Net و استفاده از مدل کانولوشنال Efficient Net B۳<sup>[۱۲]</sup> به عنوان رمزگذار آن بوده است. در بخش دوم پژوهش حاضر، به مروری کلی روش پژوهش پرداخته شده است. در بخش سوم، به الگوسازی نظری پژوهش شامل روش‌های جمع‌آوری و برچسب‌گذاری داده، توضیح جزئیات مدل استفاده شده، پارامترهای موجود در روش، راهبرد آموزش و توابع هزینه‌ی به کار رفته در مدل پرداخته شده است. در نهایت در بخش چهارم، نتایج حاصل از روش مطالعه شده استخراج شده و ارزیابی نهایی روش انجام پذیرفته است. در شکل ۱، فرایند کلی پژوهش حاضر مشاهده می‌شود.

## ۲. روش پژوهش

در بخش حاضر، به توضیح دقیق روش پژوهش پرداخته شده است. همان‌گونه در بخش مقدمه اشاره شده است، در پژوهش حاضر سعی بر این است که نواحی وجود ترک در سطوح مصالح بتنی با استفاده از الگوریتم‌های بینایی رایانه مبتنی بر یادگیری عمیق پیش‌بینی شود. همچنین محتوای دیجیتال تصاویر رنگی RGB، به صورت یک تانسور سه‌بعدی با ابعاد  $W \times H \times Ch$  قابل نمایش است که در آن،  $W$  و  $H$  بیانگر تعداد پیکسل در راستای عرض و طول تصویر و  $Ch$  نماینده‌ی سه کانال قرمز (Red)، سبز (Green) و آبی (Blue) است. هر درایه از تانسور مذکور می‌تواند مقداری بین صفر تا ۲۵۵ را اختیار کند که برهم‌نهی عمقی مقادیر تانسور اخیر در هر نقطه، نماینده‌ی یک پیکسل از تصویر است. با این تفاسیر، یک تصویر با کیفیت معمولی و با ابعاد  $۳۶۰ \times ۴۸۰$ ، ۵۱۸۴۰۰ رقم خواهد داشت که برقرار کردن ارتباط منطقی بین اعداد اشاره شده، پیچیدگی بسیار بالایی دارد. علاوه بر این، تغییر در شرایط تصویربرداری همچون تغییر شدت روشنایی یا چرخش دوربین، تغییراتی گسترده در محتوای دیجیتالی تصویر ایجاد می‌کند. پیچیدگی مسئله در حالتی مشخص می‌شود که قرار باشد براساس مقادیر عددی ذکر شده، وجود یک شیء یا جسم در داخل تصویر بررسی شود. روش‌های کلاسیک پردازش تصویر، شامل روش‌هایی همچون لبه‌یابی<sup>۱۶</sup>، اگرچه قادر به تشخیص مرزهای تصاویر بودند، با این حال، به دلیل سرعت بسیار پایین و همچنین حساسیت زیاد نسبت به شرایط تصویربرداری، عملاً با موفقیت زیادی همراه نبودند. استفاده از شبکه‌های عصبی ساده نیز با توجه به اینکه شبکه‌های عصبی، ورودی خود را به صورت برداری دریافت می‌کنند، پردازش بر روی داده‌های تصویر که ارتباط عناصر آن در دو یا سه بُعد است، موفقیت چندانی به دنبال نداشته است. اما مبنای کار شبکه‌های عصبی ساده و فرایند یادگیری ماشین، کمک شایانی به توسعه‌ی نسل جدیدتری از شبکه‌های عصبی داشته است که با عنوان شبکه‌های عصبی کانولوشنال نامیده می‌شوند. شبکه‌های عصبی کانولوشنال، برخلاف شبکه‌های عصبی ساده، به جای استفاده از واحدهای



شکل ۱. فرایند کلی استفاده شده در پژوهش حاضر.

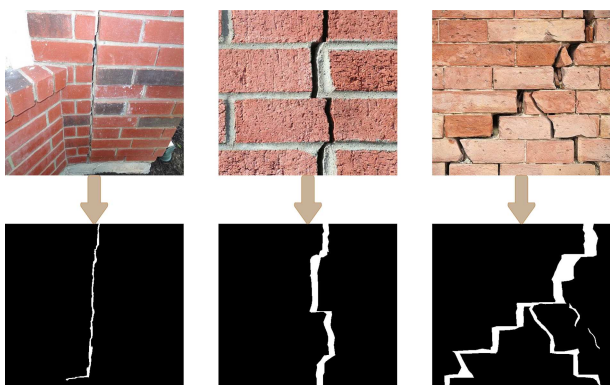


شکل ۲. مقایسه‌ی ساختار یک شبکه‌ی عصبی ساده و یک شبکه‌ی عصبی کانولوشن در فرایند دسته‌بندی.

اشیاء، علاوه بر حضور یک شی در تصویر، امکان استخراج محدوده‌ی قرارگیری آن نیز مشخص است، که معمولاً به واسطه‌ی مستطیل‌های محصورکننده قابل استخراج است که با استفاده از یک رگرسیون<sup>۱۷</sup> در کنار لایه‌ی دسته‌بندی‌کننده در انتهای مدل شبکه‌ی عصبی کانولوشن میسر می‌شود. سومین و پیشرفته‌ترین مسئله‌ی بنیادین موجود در بینایی رایانه که موضوع بحث نوشتار حاضر را نیز شامل می‌شود، بخش‌بندی تصاویر است. در بخش‌بندی تصاویر، یکایک پیکسل‌های موجود در تصویر برچسب‌گذاری می‌شوند و خروجی مدل به صورت یک تصویر تک‌کاناله با ابعاد تصویر ورودی است که مقدار پیکسل آن بیانگر دسته‌بندی است که آن پیکسل در تصویر اصلی به آن تعلق دارد. در نتیجه، خروجی روش مذکور می‌تواند جزئیات بسیار دقیقی از محل قرارگیری و ابعاد اجسام و محتوای تصویر ارائه کند. مدل‌های موجود در زمینه‌ی بخش‌بندی تصاویر، عمدتاً مدل‌هایی متقارن هستند که برخلاف سایر مدل‌های کانولوشن نیازی به لایه‌ی تماماً متصل ندارند و پس از استخراج ویژگی‌ها (مرکزگذاری)، مدل عکس فرایند پیشین را به داده اعمال (مرکزگشایی) و مقادیر مرکزگذاری شده را به تصاویر دوبعدی تک‌کاناله تبدیل می‌کنند. در شکل ۳، یک حالت کلی از یک معماری مرکزگذار - مرکزگشا که در بخش‌بندی تصاویر کاربرد دارد و در پژوهش حاضر استفاده شده است، مشاهده می‌شود که مطابق آن، این دسته از الگوریتم‌ها ابتدا با فرایند مرکزگذاری، ویژگی‌های داده را استخراج می‌کنند. فرایند مرکزگذاری با بهره‌گیری از شبکه‌های عصبی کانولوشنال و مشابه با قسمت استخراج ویژگی موجود در شکل ۲ انجام می‌پذیرد.

ستونی که مناسب کار با بردارهاست، از کرنل‌های دو و سه بُعدی استفاده می‌کنند؛ بدین ترتیب استخراج ویژگی‌ها در داده‌های دو یا چند بُعدی نیز فراهم می‌شود. در شکل ۲، مقایسه‌ی از یک شبکه‌ی عصبی ساده و یک شبکه‌ی عصبی کانولوشن برای یک مسئله‌ی طبقه‌بندی مشاهده می‌شود. ساختار کلی یک شبکه‌ی عصبی کانولوشن، مطابق شکل اخیر، از یک قسمت استخراج ویژگی تشکیل می‌شود که عملگرهای کانولوشن و توابع جمع، عمل استخراج را که معمولاً با کاهش ابعاد داده همراه است، به انجام می‌رسانند و پس از آن با تغییر فرم داده به فرم برداری یک بُعدی، امکان انتقال داده‌ی ساده شده به شبکه‌های تماماً متصل فراهم و عملیات دسته‌بندی میسر می‌شود. در یک فرایند کانولوشنال به جای واحدهای تماماً متصل در شبکه‌های عصبی، کرنل‌های دو یا سه بُعدی وارد فرایند آموزش می‌شوند و با انجام عمل کانولوشن بر روی داده‌ها، مقادیر خود را تکمیل می‌کنند. فرایند اخیر که برای طبقه‌بندی داده‌های تصویر شرح داده شد، مبنای اصلی علوم بینایی رایانه نوین است که با کمک آن می‌توان به سایر مسائل بنیادین در حوزه‌ی بینایی رایانه نیز پاسخ داد. براین اساس سه مسئله‌ی بنیادین بینایی رایانه، شامل: طبقه‌بندی، تشخیص اشیاء و بخش‌بندی قابل پوشش خواهند بود.

همچنین در یک مسئله‌ی طبقه‌بندی تصویر، خروجی مدل فقط یک برچسب است که بیانگر دسته‌ی مربوط به تصویر است. علاوه بر این، طبقه‌بندی می‌تواند وجود یا نبود یک شیء یا جسم در محتوای تصویر را پیش‌بینی کند. در تشخیص



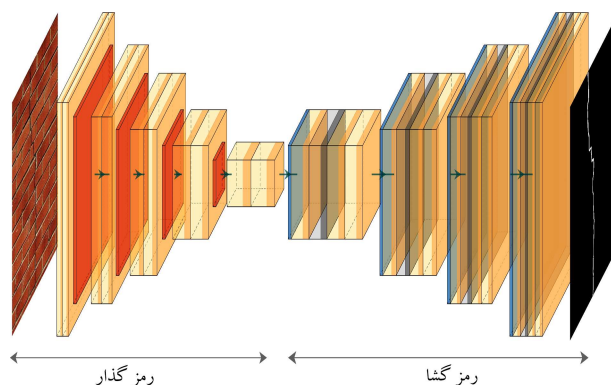
شکل ۴. نمونه‌هایی از تصاویر گردآوری شده در مجموعه‌ی داده و برچسب‌های متناظر آن‌ها.

پیکسل‌های دسته‌ی متناظر برابر یک قرار و بقیه صفر در نظر گرفته شوند؛ و در پایان، برای آموزش مدل یادگیری عمیق، ماتریس‌های تصاویر مذکور در راستای بعد سوم به یکدیگر الحاق و آماده‌ی ارسال به مدل شوند. پس از تکمیل شدن برچسب‌گذاری‌ها، داده‌ها یکدست و فرمت و ابعاد یکسانی برای آن‌ها در نظر گرفته شده است. با توجه به اینکه تصاویر ورودی همیشه کیفیت بالایی ندارند، برای تصاویر مذکور، استفاده از یک فرمت رایج و با کیفیت معمولی پیشنهاد می‌شود. برعکس تصاویر، برای برچسب‌های خروجی انتظار نتایج دقیق و با کیفیتی می‌رود؛ لذا تصاویر با فرمت JPG که یک فرمت رایج است و برچسب‌ها با فرمت PNG که دقیق‌ترین کدگذاری را در بین فرمت‌های رایج دارد، انتخاب شده است.

همچنین به عنوان یک معیار میانگین برای حفظ کیفیت و همچنین عدم افزایش بار محاسباتی، ابعاد تصاویر برابر با  $320 \times 320$  در نظر گرفته شده است. نمونه‌هایی از داده‌های ایجاد شده برای پژوهش حاضر مطابق شکل ۴ ارائه شده است. مجموعه‌ی داده‌ی ایجاد شده‌ی نهایی، شامل ۱۱۵ داده‌ی برچسب‌گذاری شده از تصاویر اعضاء مصالح بتنی از جمله آجر بوده است که ۸۱ عدد از آن‌ها برای آموزش مدل، ۱۷ عدد برای اعتبارسنجی و ۱۷ تصویر دیگر برای ارزیابی نهایی مدل استفاده شده‌اند. با توجه به محدودیت تعداد داده، از روش‌های داده‌افزایی<sup>۲۱</sup> جهت بالا بردن عملکرد مدل استفاده شده است. روش‌های داده‌افزایی، شامل: دوران تصویر، تغییر تصادفی روشنایی، اعمال نوفه‌ی گوسی که به صورت تصادفی در هر مرحله از آموزش بر روی داده‌ها اعمال شده‌اند؛ بدین ترتیب سناریوهای بیشتری در روش وارد شده و نیز احتمال وقوع بیش‌بازش نیز کاهش یافته است.

### ۲.۳. مدل یادگیری عمیق

در بخش حاضر، به شرح جزئیات مدل استفاده شده در پژوهش حاضر اشاره شده است. مدل استفاده شده در پژوهش حاضر با الهام‌گیری از U-Net<sup>[۱۱]</sup> بوده است که در آن از شبکه‌ی عصبی کانولوشنال پیشرفته‌ی EfficientNetB۳<sup>[۱۲]</sup> استفاده شده است. عنوان رمزگذار استفاده شده است. مدل رمزگذار پژوهش حاضر، چهارمین عضو از مجموعه‌ی هشت عضو مدل‌های EfficientNet است که تمامی آن‌ها مشتق شده از مدل اصلی نوشتار حاضر یعنی EfficientNetB۰ است. روش ابداعی استفاده شده برای مقیاس مدل‌ها در مجموعه‌ی اخیر با عنوان روش مقیاس مرکب<sup>۲۲</sup> امکان افزایش بیشتر عمق مدل را فراهم کرده است. در روش‌های پیشین همچون روش‌های استفاده شده در مدل‌هایی چون VGG<sup>[۱۳]</sup> مقیاس فقط در یک بعد، یعنی عمق، اعمال شده است که در نتیجه‌ی آن مشاهده می‌شود مدل VGG۱۹ با



شکل ۳. ساختار کلی یک مدل رمزگذار - رمزگشا در حل مسائل بخش‌بندی تصاویر.

پس از آن و برای رسیدن به تصویر خروجی، عمل رمزگشایی با استفاده از لایه‌های Upsampling یا Deconvolution انجام می‌شود. مدل موجود با استفاده از روش‌های یادگیری ماشین که در ادامه‌ی متن شرح داده شده است، آموزش می‌بیند و در نهایت قادر می‌شود که بر روی هر تصویر جدیدی، پیش‌بینی انجام و تصویر ترک شناسایی شده با سرعتی بالا گزارش شود. بدین ترتیب روش موجود قادر به پردازش بلادرنگ<sup>۱۸</sup> تصاویر است و می‌تواند به عنوان یک راه‌حل عملی برای مسائل واقعی استفاده شود.

### ۳. الگوسازی نظری

بخش حاضر از ۵ قسمت تشکیل می‌شود که به ترتیب به تشریح داده‌ی استفاده شده، مدل به کار گرفته شده (مدل یادگیری عمیق)، پارامترهای ارزیابی، توابع هزینه‌ی<sup>۱۹</sup> استفاده شده و فرایند آموزش مدل پرداخته شده است.

#### ۳.۱. مجموعه‌ی داده

در یادگیری ماشین و مخصوصاً یادگیری عمیق که موضوع بحث پژوهش حاضر است، داده‌ی موجود برای آموزش مدل از عناصر حیاتی به حساب می‌آید. برای ساده‌سازی بحث، اگر مدل‌های یادگیری ماشین به یک موتور درون‌سوز تشبیه شود، سوخت موتور اخیر، داده‌ی است که برای آموزش آن استفاده می‌شود. در نتیجه، در یک مدل یادگیری، عملکرد هم‌زمان دو عنصر مدل و داده، تعیین‌کننده‌ی عملکرد نهایی روش استفاده شده است. در پژوهش حاضر، با توجه به اینکه داده‌ی در دسترس و برچسب‌گذاری شده‌ی در دسترس نبوده است، در ابتدا سعی شده است تعدادی تصویر از ساختمان‌های مصالح بتنی که در آن‌ها ترک بروز کرده است، جمع‌آوری و سپس برچسب‌گذاری شوند. داده‌های استفاده شده در پژوهش حاضر، عمدتاً تصاویر موجود و قابل دسترس در اینترنت، جمع‌آوری و سپس به صورت دستی برچسب‌گذاری شده‌اند. در فرایند بخش‌بندی تصاویر که در پژوهش حاضر بررسی شده است، دقت و کیفیت برچسب‌گذاری بسیار مهم بوده است؛ چرا که در فرایند تحلیل مسئله، هر پیکسل به صورت جداگانه دسته‌بندی و در دسته‌ی مخصوص به خود قرار گرفته است. برچسب‌های خروجی ایجاد شده در فرایند برچسب‌گذاری، یک تصویر دودویی<sup>۲۰</sup> بوده است که در آن پیکسل‌های با مقدار صفر، بیانگر زمینه‌ی تصویر و پیکسل‌های با مقدار یک، نماینده‌ی ترک موجود در تصویر هستند. لازم به ذکر است اگر در فرایند برچسب‌گذاری داده‌ها، بیشتر از دو دسته در اختیار باشد، لازم است به ازاء تعداد دسته‌ها، تصویر دودویی ایجاد شود و در هر یک از آن‌ها،

جدول ۱. مقایسه‌ی عملکرد و تعداد پارامتر مشتقات EfficientNet با سایر مدل‌ها.

مدل	دقت	تعداد پارامتر (میلیون)
ResNet-۱۵۲ [۱۵]	۷۷/۸٪	۶۰
EfficientNet-B۱	۷۹/۱٪	۷/۸
ResNext-۱۰۱ [۱۶]	۸۰/۹٪	۸۴
EfficientNet-B۳	۸۱/۶٪	۱۲
NASNet-A [۱۷]	۸۲/۷٪	۸۹
EfficientNet-B۴	۸۲/۹٪	۱۹
GPipe [۱۸]	۸۴/۳٪	۵۵۶
EfficientNet-B۷	۸۴/۳٪	۶۶

جدول ۲. جزئیات مدل استفاده شده در رمزگذار (EfficientNet-B۳). [۱۷]

بلوک	سایز کرنل کانولوشن	ابعاد خروجی	تعداد لایه موجود در بلوک
کانولوشن	۳ × ۳	۱۶۰ × ۱۶۰ × ۴۰	۱
۱	۳ × ۳	۱۶۰ × ۱۶۰ × ۲۴	۲
۲	۳ × ۳	۸۰ × ۸۰ × ۳۲	۳
۳	۵ × ۵	۴۰ × ۴۰ × ۴۸	۳
۴	۳ × ۳	۲۰ × ۲۰ × ۹۶	۵
۵	۵ × ۵	۲۰ × ۲۰ × ۱۳۶	۵
۶	۵ × ۵	۱۰ × ۱۰ × ۲۳۲	۶
۷	۳ × ۳	۱۰ × ۱۰ × ۳۸۴	۲

وجود افزایش شدید در تعداد پارامترها، عملکرد چندان متفاوتی با مدل VGG ۱۶ که عمق کمتری داشته است، ارائه نداده است. مهم‌ترین ویژگی مدل EfficientNet، همزمانی دقت در کنار سرعت و سبکی مدل هاست. در جدول ۱، مقایسه‌ی مختصر از دقت دسته‌بندی برای مجموعه داده‌ی ImageNet [۱۴] و همچنین تعداد پارامترها برای زیرمجموعه‌های EfficientNet و مدل‌های هم‌تراز آن‌ها (به لحاظ دقت) و نیز عملکرد دقیق مشتقات EfficientNet در جدول ۱ ارائه شده است. مدل‌های اخیر علاوه بر اینکه دقت بسیار بالایی ارائه می‌دهند، تعداد پارامترهای بسیار کمتری را نیز شامل می‌شوند که این به معنی برتری چشمگیر در سرعت مدل‌های EfficientNet است.

در پژوهش حاضر، با توجه به نیاز همزمان به دقت و سرعت، استفاده از مدل EfficientNet-B۳ پیشنهاد شده است. برای استفاده از مدل EfficientNet-B۳ به عنوان رمزگذار لازم است لایه‌های اتصال کامل نهایی که وظیفه‌ی دسته‌بندی در مدل اصلی را دارند، حذف شوند و آخرین لایه‌ی کانولوشنال مستقیماً به رمزگشا متصل شود. مدل EfficientNet-B۳ از ۷ بلوک Mobile-inverted Bottleneck Convolution [۱۹] یا به اختصار (MBCConv) تشکیل می‌شود. مهم‌ترین ویژگی بلوک‌های اخیر در استفاده از لایه‌های کانولوشنی، عمقی تفکیک‌پذیر [۲۳] و همچنین استفاده از تابع فعال‌سازی ReLU۶ به جای ReLU است. در جدول ۲، خلاصه‌ی جزئیات مدل EfficientNet-B۳ ارائه شده است.

قسمت رمزگشای مدل EfficientNet-B۳ از قسمت رمزگشای U-Net [۲۰]

الگو می‌گیرد. با توجه به اینکه در قسمت رمزگذار، کاهش ابعاد در ۵ مرحله است؛ در قسمت رمزگشا نیز ۵ مرحله افزایش ابعاد اعمال شده است. در نتیجه رمزگشای مدل EfficientNet-B۳ شامل ۵ لایه بوده است که در هر لایه، ابعاد به دو برابر ابعاد لایه‌ی قبلی افزایش یافته‌اند. در شکل ۵، معماری کلی مدل استفاده شده مشاهده می‌شود.

### ۳.۳. آموزش مدل با الگوریتم پس‌انتشار و یادگیری انتقالی

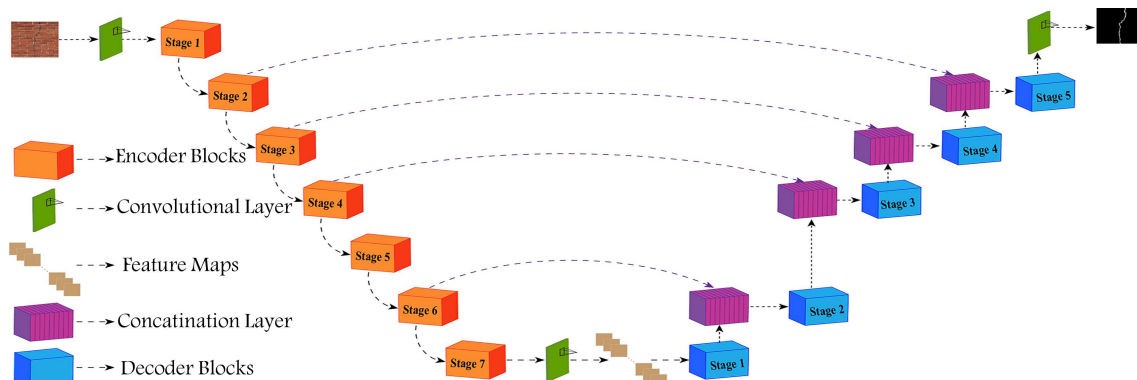
در بخش حاضر، فرایند یادگیری با ناظر [۲۴] یک مدل یادگیری عمیق بررسی شده است. فرایند کار در مسائل یادگیری ماشین و یادگیری عمیق در ساده‌ترین حالت به این صورت است که از تعدادی داده‌ی برچسب‌گذاری شده استفاده و ارتباط بین آن‌ها توسط مدل استخراج و در مدل ذخیره می‌شود؛ سپس مدل قادر است که برای داده‌های جدید و فاقد برچسب نیز پیش‌بینی انجام دهد و با دقت خوبی، برچسب متناظر با آن‌ها را گزارش کند. این فرایند به کمک الگوریتم پس‌انتشار [۲۵] که معروف‌ترین روش موجود در حوزه‌ی یادگیری ماشین است، انجام می‌پذیرد. الگوریتم مذکور شامل سه قسمت است که به این ترتیب است:

۱. انتخاب تانسوروزن‌ها ( $W$ ) در ابعادی که بتواند در رابطه‌ی  $x, y = Wx$  را به  $y$  متناظر کند و سپس پر کردن درایه‌های آن با مقادیر تصادفی (Random).
۲. محاسبه‌ی مقدار  $\hat{y}$  از رابطه‌ی  $\hat{y} = Wx$  و محاسبه‌ی اختلاف  $y$  و  $\hat{y}$  از طریق یکی از توابع هزینه (Loss Function)؛
۳. اصلاح مقادیر  $W$  با استفاده از الگوریتم بهینه‌ساز به طوری که مقدار هزینه (Loss)، کمینه (minimize) شود و بازگشت به مرحله‌ی ۲ و تکرار این فرایند تا جایی که  $W$  بهینه به دست آید.

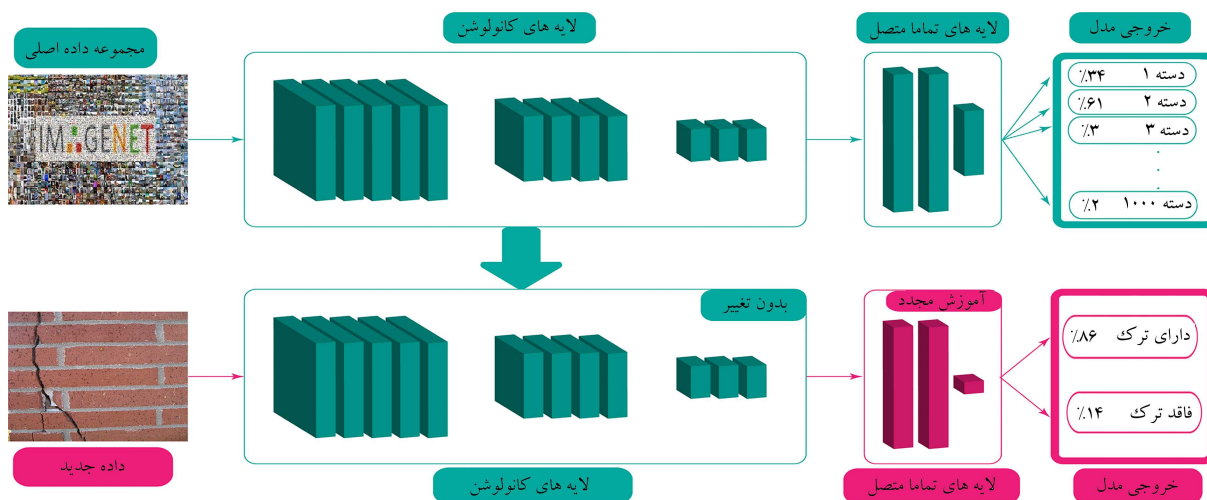
در فرایند یادگیری انتقالی استفاده شده در پژوهش حاضر، مدل از حالت صفر شروع به آموزش نکرده و به جای پر کردن  $W$  با مقادیر تصادفی، مقادیر حاصل از آموزش مدل برای مجموعه داده‌ی Image-Net [۱۴] استفاده شده است. با توجه به اینکه مجموعه داده‌ی مذکور، یکی از کامل‌ترین و جامع‌ترین مجموعه‌ی داده‌ها در حوزه‌ی بینایی رایانه است، مقادیر اولیه‌ی اشاره شده توانایی بالایی در استخراج ویژگی‌های [۲۶] تصاویر داشته و فرایند آموزش بیشتر در لایه‌های آخر که وظایفی چون: دسته‌بندی، رگرسیون یا تقسیم‌بندی را به عهده دارند، متمرکز شده است. در شکل ۶، نمونه‌ی یکی از یک فرایند یادگیری انتقالی برای آموزش یک مدل دسته‌بندی ترک مشاهده می‌شود. در پژوهش حاضر، با توجه به اینکه مسئله از نوع بخش‌بندی است، در لایه‌هایی که استخراج ویژگی انجام شده است (لایه‌های رمزگذار)، از یادگیری انتقالی و وزن‌های از پیش آموزش‌دیده استفاده شده و تغییر زیادی در لایه‌های مذکور رخ نداده است؛ اما در قسمت رمزگشا، استفاده از یادگیری انتقالی میسر نبوده و لازم بوده است که فرایند آموزش کاملاً انجام شود.

### ۴.۳. پارامترهای ارزیابی

برای سنجش و ارزیابی نتایج حاصل از پژوهش حاضر نیاز به پارامترهایی است که بتوانند تفسیر درستی از کیفیت فرایند بخش‌بندی تصویر ارائه دهند. به منظور به دست آوردن دقت برای هر یک از داده‌ها، لازم بود که شباهت مقادیر موجود در برچسب اولیه با برچسب خروجی از مدل سنجیده شود. با توجه به عدم توازن شدید موجود در تصاویر ترک بررسی شده، بهتر بوده از پارامترهایی استفاده شود که دقت هر دسته را به صورت نسبی محاسبه کنند. پارامتر اصلی که در مطالعه‌ی حاضر بررسی شده است، معیار  $F1$  [۲۷] است که نوعی توزیع یکنواخت از دو پارامتر صحت [۲۸] و فراخوانی [۲۹]



شکل ۵. معماری کلی مدل استفاده شده در پژوهش حاضر.



شکل ۶. یادگیری انتقالی برای فرایند طبقه‌بندی تصویر.

در تلاش برای کمیته کردن مثبت‌های کاذب، تمایل خواهد داشت که پیکسل‌های بیشتری را به عنوان منفی دسته‌بندی کند؛ در نتیجه، در این بین تعداد زیادی منفی کاذب ایجاد می‌شود که کاهش فراخوانی را در پی خواهد داشت. عکس این مطلب نیز صادق است. اگر فقط بر بهینه‌سازی فراخوانی تمرکز شود، مدل به هر قیمتی سعی در کمیته‌سازی منفی‌های کاذب خواهد داشت. دقت شود که در بدترین حالت، مدل می‌تواند تمامی پیکسل‌ها را به عنوان ترک دسته‌بندی کند و به این ترتیب  $FN$  برابر صفر می‌شود و همچنین با توجه به صفر نبودن  $TP$ ، فراخوانی برابر  $100\%$  خواهد شد که با مطلوب مسئله پژوهش حاضر، فاصله‌ی بسیار زیادی دارد. اما ترکیب همزمان دو پارامتر مذکور می‌تواند بهترین نتایج را به همراه آورد، چرا که مکمل همدیگر هستند و نقاط ضعف یکدیگر را نیز کاملاً پوشش می‌دهند. بر این اساس معیار دقت  $F1$  به صورت توزیعی یکنواخت از دو پارامتر پیشین قابل تعریف است. پارامتر  $F1$  مطابق رابطه‌ی ۳ تعریف می‌شود:

$$F1 = \frac{2 \times Pr \times Re}{Pr + Re} \quad (3)$$

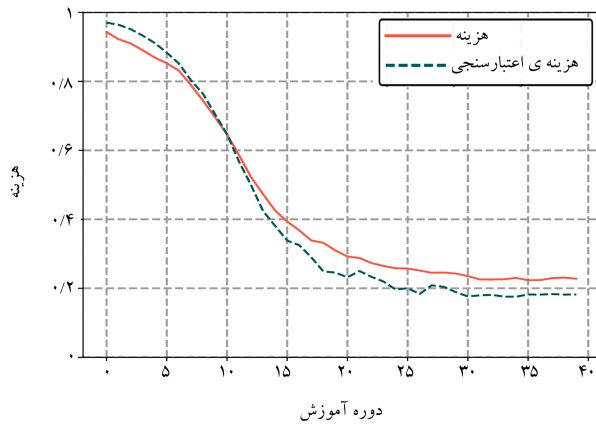
پارامتر  $F1$  به طور همزمان شامل صحت و فراخوانی می‌شود. معیار  $F1$  یکی از محبوب‌ترین و پرکاربردترین پارامترها در سنجش دقت مدل‌های بخش‌بندی تصاویر است. در پژوهش حاضر، درنهایت هدف رسیدن به بیشترین مقدار معیار  $F1$  بوده است.

است. به هر میزان که مقدار صحت بیشتر باشد، احتمال پیش‌بینی پیکسل غلط در مدل کمتر خواهد بود؛ در حالی که بالا رفتن مقدار فراخوانی به تشخیص تمام و کمال پیکسل‌های ترک کمک خواهد کرد. واضح است که برای رسیدن به پاسخ بهینه لازم است که دو پارامتر اخیر به‌طور همزمان پیشرفت کنند، چرا که بهینه‌شدن یک مورد آن‌ها به تنهایی نمی‌تواند نتایج مطلوبی حاصل کند. به عبارتی دیگر، اگر در یک پیش‌بینی، صحت پیش‌بینی زیاد باشد، اما فراخوانی کم باشد، احتمال از دست رفتن تعداد زیادی از پیکسل‌های ترک وجود خواهد داشت؛ در حالی که اگر عکس موضوع اتفاق افتد، تعداد زیادی از پیکسل‌ها خواهند بود که به اشتباه به عنوان ترک دسته‌بندی می‌شوند. برای روشن شدن موضوع می‌توان به روابط ۱ و ۲ دقت کرد:

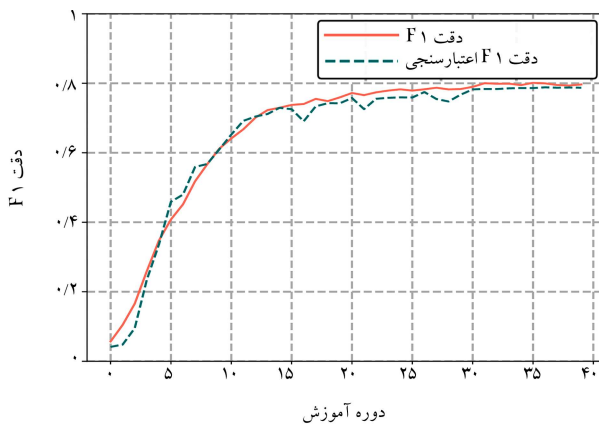
$$Pr = \frac{TP}{TP + FP} \quad (1)$$

$$Re = \frac{TP}{TP + FN} \quad (2)$$

که در آن‌ها،  $TP$  بیانگر تعداد پیکسل‌هایی است که به درستی ترک تشخیص داده می‌شوند (مثبت صحیح).  $FP$  نیز بیانگر تعداد پیکسل‌هایی است که به اشتباه ترک تشخیص داده می‌شوند (مثبت کاذب) و  $FN$  بیانگر تعداد پیکسل‌هایی است که به عنوان زمینه پیش‌بینی می‌شوند، در حالی که این پیش‌بینی نادرست است (منفی کاذب). قابل درک است که اگر صحت به تنهایی بهینه شود، مدل



شکل ۷. نمودار تغییرات هزینه در طی آموزش.



شکل ۸. نمودار تغییرات دقت F1 در طی آموزش.

شکل های ۷ و ۸ مشاهده می شوند. مطابق شکل های ۷ و ۸، در حالت کلی، مقادیر پارامترهای مختلف برای مدل در حالت آموزش و اعتبارسنجی باید به صورت هماهنگ تغییر کنند و اگر خلاف آن رخ دهد، بیانگر وقوع بیش برآزش در مدل است.

همچنین مطابق جدول ۳، در هنگامی که مدل به بهترین حالت آموزش رسیده باشد، پارامترهای اعتبارسنجی تقریباً ثابت مانده و در محدوده بی بسیار کوچک نوسان داشته اند؛ این موضوع از مرحله ۳۵ به بعد به خوبی مشاهده می شود. لازم است به ذکر است اگر فرایند آموزش برای مدل مذکور در چند مرحله دیگر ادامه می یافت، مقادیر پارامترهای اعتبارسنجی دچار افت ناگهانی می شد، در حالی که پارامترهای آموزش همچنان به پیشرفت ادامه می دادند. همان طور که پیش از این اشاره شد، به پدیده ی اخیر، بیش برآزش گفته می شود که در یک مدل یادگیری ماشین، فرایند مطلوبی نیست و بیانگر افت عملکرد مدل در برابر داده های جدید است. پس از اینکه با توجه به داده های موجود در جدول ۳ اطمینان حاصل شد که مدل به بیشترین مقدار دقت خود رسیده است، لازم است که مدل با استفاده از داده ی ارزیابی نهایی، ارزیابی و دقت نهایی مدل مشخص شود. با انجام ارزیابی برای داده ی مذکور، مقادیر جدول ۴ برای پارامترهای عنوان شده در قسمت دوم به دست آمده است.

با توجه به جدول ۴ می توان دریافت که مقادیر دقت برای داده ی ارزیابی، شباهت بسیاری با مقادیر پارامترهای اعتبارسنجی موجود در جدول ۳ دارد. بدین ترتیب می توان نسبت به صحت و دقت مدل مطالعه شده اطمینان حاصل کرد و

### ۵.۳. تابع هزینه

با توجه به پارامترهای تعریف شده در مرحله ی پیشین، لازم است تابعی تعریف شود که در حین آموزش مدل یادگیری عمیق، مقدار پارامتر موردنظر به بیشترین مقدار ممکن برسد. به طور کلی، در طی یک فرایند آموزش، در هر مرحله از آموزش لازم است که اختلاف بین برچسب های اولیه و خروجی از مدل مشخص شود. این اختلاف که به تابع هزینه  $\mathcal{L}$  موسوم است، باید ارتباط و نزدیکی لازم با پارامتر ارزیابی مسئله را داشته باشد. همان طور در بخش قبل گفته شد، هدف در پژوهش حاضر دستیابی به بیشترین مقدار برای معیار سنجش  $F1$  است؛ این امکان وجود دارد که تابع هزینه به صورتی تعریف شود که معیار  $F1$  را مستقیماً بهینه سازد. تابع هزینه ی مذکور که Dice-Loss نام دارد، مستقیماً مسئله را برای رسیدن به بیشترین مقدار  $F1$  بهینه می سازد. تابع هزینه ی اخیر در حقیقت تفاضل معیار  $F1$  موجود در مدل را از مقدار پیشینه یعنی ۱ بیان می کند.

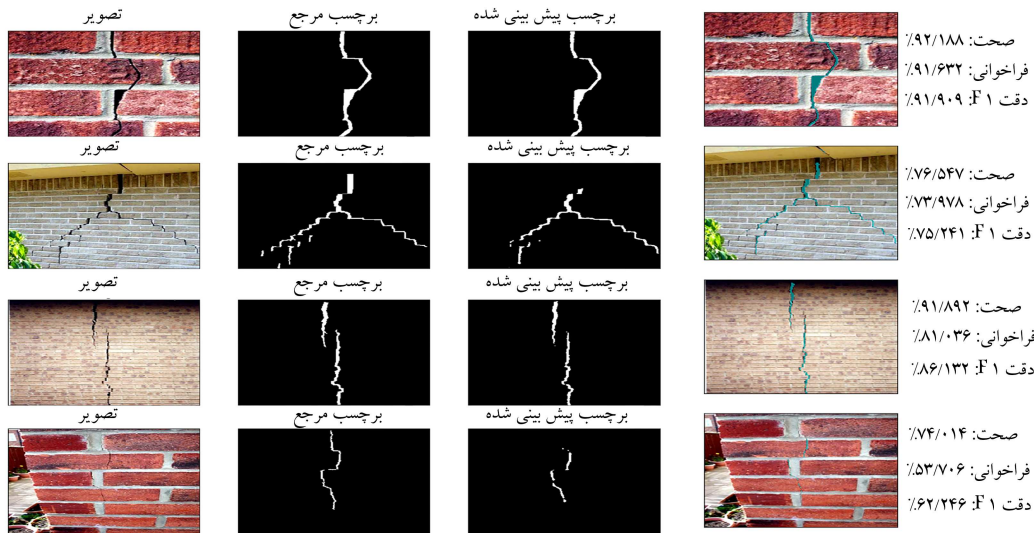
### ۴. نتایج و بحث

پس از آنکه پارامترهای اشاره شده در مرحله ی قبل به مدل یادگیری عمیق اعمال شد، فرایند آموزش مدل آغاز شده است. در مطالعه ی حاضر، مدل برای ۴۰ مرحله آموزش داده شد و در پایان هر مرحله، فرایند اعتبارسنجی انجام پذیرفت. دقت شود که با توجه تعداد کم داده ی آموزش و عمق نسبتاً زیادی مدل، احتمال وقوع پدیده ی بیش برآزش در مدل بسیار بالاست و در این حالت اعتبارسنجی مرحله ی کمک خواهد کرد که تغییرات مدل رصد شود و بدین ترتیب می توان در لحظه ی بیش برآزش مدل، از وقوع پدیده ی اخیر جلوگیری کرد. لازم به ذکر است بیش برآزش به حالتی گفته می شود که فرایند آموزش بیشتر از حد ضرورت برای مدل انجام پذیرد؛ در این حالت دقت پیش بینی مدل برای داده های آموزش نسبتاً زیاد می شود، اما در مقابل، پیش بینی مدل برای داده های جدید بسیار کم خواهد بود. در فرایند اعتبارسنجی مرحله ی آموزش طی چند مرحله انجام می پذیرد و در پایان هر یک از مراحل مذکور، یک پیش بینی برای مجموعه ی داده، اعتبارسنجی انجام می شود. چنانچه نتایج پیش بینی اشاره شده انطباق مناسبی با دقت آموزش داشته باشد، مدل در ناحیه ی امن است، اما زمانی که با وجود کاهش مقدار هزینه و افزایش دقت برای داده ی آموزش، دقت پیش بینی برای داده ی اعتبارسنجی شروع به کاهش کرد، بیش برآزش مدل آغاز می شود و باید فرایند آموزش متوقف شود. در جدول ۳، جزئیات فرایند آموزش و اعتبارسنجی مدل برای ۴۰ مرحله ی ذکر شده ارائه شده است. فرایند آموزش با استفاده از زبان برنامه نویسی پایتون و نیز کتابخانه ی Tensorflow<sup>[۲۰]</sup> که یکی از شناخته شده ترین پلتفرم های هوش مصنوعی است، انجام پذیرفته است. همچنین لازم به ذکر است به دلیل پیچیدگی بسیار زیاد آموزش مدل های یادگیری عمیق و هزینه ی محاسباتی بسیار بالای آنها، آموزش مدل مذکور با استفاده پردازنده ی CPU میسر نبوده و آموزش با بهره گیری از یک پردازنده ی گرافیکال از نوع N-Vidia Tesla K80 انجام پذیرفته است. دلیل اصلی استفاده از پردازنده های گرافیکی در فرایند یادگیری عمیق آن است که پردازنده های مذکور به واسطه ی تعداد بسیار زیاد هسته می توانند نوعی پردازش موازی در واحدهای کوچک را به هنگام آموزش ایجاد کنند که این موضوع باعث افزایش چشمگیر سرعت پردازش در پردازنده های اخیر نسبت به CPU می شود. نمودارهای تغییرات هزینه و دقت مدل در طی فرایند آموزش در

جدول ۳. جزئیات پارامترهای مدل در طی آموزش.

مرحله	هزینه	دقت F1	هزینه ی اعتبارسنجی	دقت F1 اعتبارسنجی
۱	./۹۴۳۲۱۱	./۰۵۶۹۶۱	./۹۷۰۳۳۲	./۰۴۱۳۷۵
۲	./۹۲۲۳۳	./۱۰۴۶۲۵	./۹۶۴۱۲۶	./۰۴۷۹۹۶
۳	./۹۱۰۱۳۵	./۱۶۵۶۳۵	./۹۵۱۵۳۷	./۰۹۵۰۷۹
۴	./۸۹۰۴۷۴	./۲۵۹۱۳۷	./۹۳۲۸۴	./۲۳۴۳۰۹
۵	./۸۶۹۲۸۳	./۳۴۷۰۵۶	./۹۱۱۳۴۸	./۳۳۵۳۶۴
۶	./۸۵۲۶۱۹	./۴۰۸۱۰۸	./۸۸۳۴۴	./۴۵۹۰۰۸
۷	./۸۳۱۳۶۵	./۴۵۱۶۰۳	./۸۵۲۰۲۶	./۴۷۹۷۳۵
۸	./۷۸۹۷۰۸	./۵۱۸۰۳۴	./۸۰۳۷۴۶	./۵۵۹۲۱۷
۹	./۷۴۴۳۷۷	./۵۶۶۶۶۲	./۷۶۴۵۶۷	./۵۶۷۷۰۳
۱۰	./۶۹۶۵۷۶	./۶۱۱۶۲	./۷۰۵۲۴۱	./۶۰۹۴۲۱
۱۱	./۶۴۶۴۷۷	./۶۴۱۳۰۷	./۶۴۷۲۳۱	./۶۵۲۳۸۹
۱۲	./۵۸۶۱۹۵	./۶۶۷۶۸۳	./۵۶۷۴۲۵	./۶۹۱۹۶۷
۱۳	./۵۲۲۵۱۵	./۷۰۱۰۸۲	./۴۹۸۸۹۸	./۷۰۴۲۹۸
۱۴	./۴۷۲۶۷۷	./۷۲۲۷۹۲	./۴۲۳۹۴۴	./۷۱۱۵۷۶
۱۵	./۴۲۵۶۱۹	./۷۲۹۷۰۳	./۳۸۰۹۳۴	./۷۲۹۴۴۶
۱۶	./۳۹۳۴۷۹	./۷۳۷۸۶۲	./۳۳۸۴۰۸	./۷۲۵۴۲۱
۱۷	./۳۶۷۹۹۲	./۷۴۰۴	./۳۲۵۰۱۴	./۶۹۰۸۱۶
۱۸	./۳۳۸۷۱۵	./۷۵۵۱۴۵	./۲۸۹۹۹۶	./۷۳۳۷۲۴
۱۹	./۳۳۱۹۱۸	./۷۴۸۷۳۳	./۲۴۹۰۸	./۷۴۲۷۴۴
۲۰	./۳۰۹۷۴۸	./۷۵۹۵۹۲	./۲۴۵۶۱۳	./۷۴۲۵۹۶
۲۱	./۲۹۲۳۲۹	./۷۷۲۱	./۲۳۱۹۴۲	./۷۵۷۶۷۶
۲۲	./۲۸۷۸۳۶	./۷۶۵۹۴۸	./۲۵۰۱۴۵	./۷۲۵۶۶۲
۲۳	./۲۷۳۳۱۳	./۷۷۳۹۳۹	./۲۳۳۲۰۳	./۷۵۵۴۴۳
۲۴	./۲۶۴۸۰۴	./۷۷۸۹۸۳	./۲۱۹۶۴۸	./۷۵۸۲۲۷
۲۵	./۲۵۸۱۳۷	./۷۸۲۶۱۵	./۱۹۶۴۳۶	./۷۵۹۴۷۷
۲۶	./۲۵۷۳۷۸	./۷۷۹۰۱۲	./۲۰۰۳۰۸	./۷۵۹۱۶۳
۲۷	./۲۵۱۴۹۲	./۷۸۲۷۷۲	./۱۸۳۵۸۴	./۷۷۴۸۰۲
۲۸	./۲۴۵۱۵۱	./۷۸۷۲۳۵	./۲۰۸۷۱۷	./۷۵۴۷۳
۲۹	./۲۴۵۷۵۳	./۷۸۲۴۲۲	./۲۰۴۳۱۴	./۷۴۷۶۷۶
۳۰	./۲۴۲۸۰۸	./۷۸۳۴۷۶	./۱۸۸۹۰۱	./۷۶۷۱۳۳
۳۱	./۲۳۵۵۴۲	./۷۸۹۷۹۶	./۱۷۶۴۶۳	./۷۸۲۴۵
۳۲	./۲۲۵۷۸۸	./۸۰۰۲۶۱	./۱۷۹۶۳۴	./۷۸۳۶۷۸
۳۳	./۲۲۵۵۶۴	./۷۹۸۶۹۸	./۱۸۱۰۴۸	./۷۸۳۵۱۴
۳۴	./۲۲۶۲۱۱	./۷۹۸۹۹	./۱۷۵۹۶۱	./۷۸۵۵۲۹
۳۵	./۲۳۰۱۲۴	./۷۹۴۸۰۴	./۱۷۵۷۵۱	./۷۸۶۴۰۷
۳۶	./۲۲۳۴۶۵	./۸۰۱۳۰۴	./۱۸۱۹۵۵	./۷۸۶۲۷
۳۷	./۲۲۳۹۲	./۷۹۹۸۳۱	./۱۸۱۸۲۲	./۷۸۸۴۱۸
۳۸	./۲۲۹۵۳۵	./۷۹۵۰۷۴	./۱۸۳۲۰۷	./۷۸۷۲۳۳
۳۹	./۲۳۰۹۶۷	./۷۹۳۶۲۲	./۱۸۱۸۴۹	./۷۸۷۸۶۵
۴۰	./۲۲۷۷۴۹	./۷۹۶۶۶۶	./۱۸۲۰۱۲	./۷۸۷۱۹۸





شکل ۹. خروجی مدل برای تعدادی از داده‌های ارزشیابی.

جدول ۴. مقادیر پارامترهای ارزیابی نهایی مدل.

پارامتر	مقدار
هزینه	۰/۱۲۷۴۷
صحت	۰/۸۱۴۴۴
فراخوانی	۰/۷۱۴۱۱
دقت F1	۰/۷۵۳۶۶

از آن برای پیش‌بینی و شناسایی ترک برای تصاویر و سناریوهای مشابه استفاده کرد. خروجی مدل برای تعدادی از تصاویر در شکل ۹ مشاهده می‌شود.

## ۵. نتیجه‌گیری

در نوشتار حاضر، با استفاده از هوش مصنوعی و الگوریتم‌های یادگیری عمیق، اقدام به شناسایی ترک در اعضاء مصالح بنایی شده است. این فرایند با استفاده از روش بخش‌بندی تصاویر انجام شده است که طی آن یکایک پیکسل‌های موجود در تصویر به دو دسته‌ی ترک و غیرترک دسته‌بندی و بدین ترتیب نواحی وجود ترک در تصویر با بیشترین دقت ممکن استخراج شده‌اند. فرایند یادگیری استفاده شده از نوع یادگیری عمیق با ناظر بوده است که طی آن از داده‌های برچسب‌گذاری شده برای آموزش مدل استفاده شده و پس از تکمیل فرایند آموزش، مدل امکان پیش‌بینی برای داده‌های جدید را پیدا کرده است. تعداد ۱۱۵ تصویر از سازه‌های مصالح بنایی که دچار ترک شده‌اند، به‌وسیله‌ی اینترنت تهیه شدند و پس از برچسب‌گذاری دستی آن‌ها، فرایند آموزش مدل صورت پذیرفت. یادگیری مدل در طی ۴۰ مرحله با اعتبارسنجی مرحله‌ی و همچنین استفاده از روش یادگیری انتقالی انجام شده است. لازم به ذکر است در

جمع‌آوری داده‌های استفاده شده در پژوهش حاضر سعی شده است از تصاویری با زمینه‌های نسبتاً پیچیده‌تر نیز استفاده شود؛ در نتیجه داده‌ها به تصاویر واقعی نزدیک‌تر هستند و مدل نهایی کاربردی‌تر می‌شود. به عبارت دیگر، در صورت استفاده از تصاویر دارای زمینه‌ی ساده، دقت اسمی مدل بیشتر می‌شود؛ اما در سناریوهای واقعی عملکرد نسبتاً ضعیفی ارائه خواهند داد. استفاده از روش‌های داده‌افزایی از دیگر روش‌های بهینه‌سازی مدل بوده است که در پژوهش حاضر استفاده شده است. مدل استفاده شده در پژوهش حاضر یک الگوریتم رمزگذار - رمزگشا U شکل

بوده است که از مدل شبکه‌ی عصبی کانولوشنال پیشرفته‌ی EfficientNet-B3 به عنوان رمزگذار آن استفاده و پس از آن طی ۵ مرحله‌ی رمزگشایی، خروجی مدل حاصل شده است. در پایان، مدل موجود دقتی برابر ۷۵/۳۶۶٪ ارائه داده است. که با توجه به پیچیدگی‌های داده‌ی موردنظر، عملکرد مناسب و قابل قبولی است. اگرچه دقت به‌دست آمده در مدل حاضر به عنوان یک مدل بینایی رایانه مقدار قابل قبولی دارد، اما این عملکرد امکان بهبود نیز دارد که می‌تواند به دو روش انجام پذیرد و مقدمه‌ی برای مطالعات پیش‌رو باشد: نخست جمع‌آوری مقدار بسیار بیشتر داده به طوری که تمامی سناریوهای موجود در محیط‌های مرتبط وارد مجموعه شوند و همچنین کمیت داده‌ها آنقدر زیاد باشد که مدل به دقت بیشتری دست پیدا کند و راه حل دوم، اینکه فرایند تشخیص ترک به دو مرحله شکسته شود، نخست تشخیص و استخراج قسمتی از تصویر که مصالح موردنظر در آن قرار گرفته و سپس تشخیص ترک بر روی تصویر کوچک‌تری که در گام قبل استخراج شده است. در حالت دوم، عملاً دو مدل استفاده شده‌اند که هر کدام یک گام را به انجام می‌رسانند. علاوه بر این، تمرکز مطالعه‌ی حاضر برای تشخیص ترک بوده است، در حالی که در مبحث پایش سلامت سازه‌ها، طیف گسترده‌ی آسیب‌ها برای ساختمان‌ها تعریف شده است که ترک فقط یکی از آن‌هاست و به مابقی آسیب‌ها می‌توان در مطالعات آتی توجه کرد.

## پانویس‌ها

1. deep learning

- convolutional neural networks
- Krizhevsky
- classification
- ImageNet

6. Cha
7. transfer learning
8. Feng
9. structural health monitoring
10. object detection
11. semantic segmentation
12. fully convolutional networks
13. Yang
14. encoder - decoder
15. Liu
16. edge detection
17. regressor
18. real-time
19. loss function
20. binary image
21. data augmentation
22. compound scaling method
23. Depth-wise separable convolutions
24. supervised learning
25. Back-propagation algorithm
26. feature extraction
27. F1-score
28. precision
29. recall
30. true positive
31. false positive
32. false negative
33. loss function

### منابع (References)

1. Krizhevsky, A., Sutskever, I. and Hinton, G.E. "Imagenet classification with deep convolutional neural networks", *Advances in Neural Information Processing Systems*, **25**(2), pp. 1097-1105 (2012).
2. O. Russakovsky, J. Deng, H. Su. and et al. "Imagenet large scale visual recognition challenge", *International Journal of Computer Vision*, **115**(3)pp.211-252 (2015).
3. Cha, Y.J, Choi, W. and Buyukozturk, O. "Deep learning-based crack damage detection using convolutional neural networks", *Computer-Aided Civil and Infrastructure Engineering*, **32**(5), pp. 361-378 (2017).
4. Cha, Y.J, Choi, W. and Suh, S. "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types", *Computer-Aided Civil and Infrastructure Engineering*, **33**(9), pp. 731-747 (2018).
5. Mohtasham Khani, M., Vahidnia, S., Ghasemzadeh, L. and et al. "Deep-learning-based crack detection with applications for the structural health monitoring of gas turbines", *Structural Health Monitoring*, **19**(5), pp. 1440-1452 (2020).
6. Feng, C., Zhang, H., Wang, S. and et al. "Structural damage detection using deep convolutional neural network and transfer learning", *KSCE Journal of Civil Engineering*, **23**(10), pp. 4493-4502 (2019).
7. Long, J., Shelhamer, E. and Darrell, T. "Fully convolutional networks for semantic segmentation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440 (2015).
8. Yang, X., Li, H., Yu, Y. and et al. "Automatic pixel-level crack detection and measurement using fully convolutional network", *Computer-Aided Civil and Infrastructure Engineering*, **33**(12), pp. 1090-1109 (2018).
9. Zhang, L., Yang, F., Zhang, Y.D. and et al. "Road crack detection using deep convolutional neural network", *2016 IEEE International Conference on Image Processing (ICIP)*, *IEEE*, pp. 3708-3712 (2016).
10. Ronneberger, O., Fischer, P. and Brox, T. "U-net: Convolutional networks for biomedical image segmentation", *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 234-241 (2015).
11. Liu, Z., Cao, Y., Wang, Y. and et al. "Computer vision-based concrete crack detection using U-net fully convolutional networks", *Automation in Construction*, **104**, pp. 129-139 (2019).
12. Tan, M. and Le, Q. "Efficientnet: Rethinking model scaling for convolutional neural networks", *International Conference on Machine Learning, PMLR*, pp. 6105-6114 (2019).
13. Simonyan, K. and Zisserman, A. "Very deep convolutional networks for large-scale image recognition", *ArXiv Preprint ArXiv*, **1409**, pp. 1556 (2014).
14. Deng, J., Dong, W., Socher, R. and et al. "Imagenet: A large-scale hierarchical image database", *2009 IEEE Conference on Computer Vision and Pattern Recognition, Ieee*, pp. 248-255 (2009).
15. He, K., Zhang, X., Ren, S. and et al. "Deep residual learning for image recognition", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778 (2016).
16. Xie, S., Girshick, R. and Dollar, Z. and et al. "Aggregated residual transformations for deep neural networks", *Proceedings of the IEEE Conference on Computer Vision and pattern recognition*, pp. 1492-1500 (2017).
17. Zoph, B., Vasudevan, V., Shlens, J. and et al. "Learning transferable architectures for scalable image recognition", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8697-8710 (2018).
18. Huang, Y., Cheng, Y., Bapna, A. and et al. "Gpipe: Efficient training of giant neural networks using pipeline parallelism", *Advances in neural information Processing Systems*, **32**, pp. 103-112 (2019).
19. Sandler, M., Howard, A., Zhu, M. and et al. "Mobilenetv2: Inverted residuals and linear bottlenecks", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4510-4520 (2018).
20. Abadi, M., Agarwal, A., and Barham, P. "TensorFlow: Large-scale machine learning on heterogeneous Systems", *Journal article, CoRR* (2015).