روش پیشنهادی شامل یک مرحله پیش پردازش و به دنبال آن MaxNet و یک فرآیند محاسبه شاخص شباهت است که در شکل3\_2 نشان داده شده است.



شکل 3\_\_2 بلوک دیاگرام روش پیشنهادی

 این روش شامل دو مرحله اصلی است، یعنی مرحله طبقه بندی و مرحله بازیابی. مرحله طبقه بندی بیشتر برای آموزش مدل MaxNet استفاده می شود. در حالی که برای آزمایش سیستم CBIR پیشنهادی، از فاز بازیابی استفاده می شود. تصاویر موجود در مجموعه داده به مقیاس خاکستری تبدیل می شوند که به عنوان کانال چهارم در کنار کانال های قرمز، سبز و آبی اضافه می شود تا یک ورودی چهار بعدی برای مدل MaxNet تشکیل دهد. تصاویر برای حفظ مقدار میانگین صفر و واریانس نزدیک به یک نرمال می شوند. یک کانال تصویر نرمال شده با کم کردن مقدار میانگین از مقادیر پیکسل و تقسیم مقدار پیکسل حاصل بر انحراف استاندارد ایجاد می شود. می توان آن را به صورت زیر بیان کرد:

$I\_{n}$=$\frac{( I- µ)}{σ}$

در جایی که In، I، μ و σ به ترتیب نشان دهنده کانال نرمال شده، کانال اصلی، مقدار شدت میانگین و انحراف استاندارد هستند.

برای کاهش استفاده از منابع سیستم، تصاویر به وضوح یکنواخت 100 × 100 پیکسل تغییر شکل می دهند. استفاده از تصاویر با وضوح پایین برای آموزش مدل MaxNet، کارایی سیستم را بدون در نظر گرفتن دقت بازیابی تصویر افزایش می‌دهد، زیرا هسته‌های کمتری برای جمع شدن روی نقشه‌های ویژگی ورودی مورد نیاز است. این تصاویر برای اهداف آموزشی از مدل MaxNet عبور داده می شوند و در نهایت بردارهای ویژگی همه تصاویر استخراج می شوند تا پایگاه داده ویژگی ها را تشکیل دهند. این پایگاه داده هنگام بازیابی تصاویر مربوط به تصویر پرس و جو استفاده می شود. روش شناسی تفصیلی در بخش های فرعی بعدی مورد بحث قرار می گیرد.

3\_4 شبکه های عصبی کانولوشنال

 یک سیستم بازیابی تصویر مبتنی بر یادگیری عمیق در این کار پیشنهاد شده است. مدل DCCN بر جنبه های متعدد موجود در یک تصویر برای استخراج اطلاعات جهانی و همچنین محلی تمرکز می کند. مدل‌های DCNN بردارهای ویژگی را مستقیماً از تصاویر می‌آموزند، برخلاف روش‌های سنتی استخراج ویژگی دست‌ساز .این امر DCNN را علیرغم تئوری ریاضی نسبتاً ساده اش به یک ابزار بینایی رایانه ای قدرتمند تبدیل می کند. هدف DCNN یادگیری هسته‌های مختلف است که بهترین نمایش انتزاعی از ویژگی‌ها را در مجموعه داده ایجاد می‌کنند. نقشه ویژگی پیچیدگی Fci به صورت زیر محاسبه می شود.

$F\_{ci}$=$b\_{c}$+$\sum\_{i}^{}M\_{ci}$\*$I\_{i}$

که در آن، هسته‌های کانولوشن و نقشه ورودی به ترتیب با bc،Mci و Ii نشان داده می‌شوند.

 عملگر ∗ برای نشان دادن عملیات پیچیدگی استفاده می شود. مدل‌های DCCN عموماً با چیدن سه لایه اصلی بر روی یکدیگر ساخته می‌شوند. این لایه‌ها شامل لایه‌های کانولوشن، لایه‌های pooling و لایه‌های کاملاً متصل هستند. این لایه ها بین لایه های ورودی و خروجی قرار می گیرند. به غیر از این سه لایه، بسیاری از لایه‌های فرعی مانند حذف تصادفی، نرمال‌سازی دسته‌ای، لایه‌های جمع‌کننده و تفریق‌کننده در دسترس هستند. لایه پیچیدگی مهمترین واحد CNN را تشکیل می دهد. این لایه ای است که حاوی هسته های با اندازه های مختلف است که بر روی نقشه ورودی به هم می پیوندند تا ویژگی های مهم موجود در تصویر را به دست آورند. لایه های ادغام به طور گسترده برای کاربردشان در نمونه برداری پایین از نقشه ویژگی ورودی و در برخورد با تغییر ناپذیری ترجمه استفاده می شود. لایه Max-Pooling مورد استفاده در این تنظیم، حداکثر مقدار ویژگی را در زیربخش مستطیلی به اندازه p × p نقشه ویژگی انتخاب می کند. لایه ادغام به پایین نقشه ویژگی ورودی را با افزایش اندازه پنجره مستطیلی به نام اندازه استخر یا با افزایش گام استخر نمونه برداری می کند. گام استخر تعیین می‌کند که یک پنجره جهشی روی نقشه ویژگی ورودی چقدر بزرگ است. پنجره Max-pool با اندازه p × p خروجی P(i,j) را برای فعال سازی در نقشه ویژگی F به عنوان محاسبه می کند.

$P\_{(i‚j)}$= max($F\_{(i+p‚j+p)}$

لایه های کاملاً متصل معمولاً در نزدیکی انتهای معماری CNN برای خروجی احتمالات توزیع کلاس استفاده می شوند. مدل MaxNet پیشنهادی از این لایه‌ها برای تشکیل یک چارچوب عمومی برای کار بازیابی تصویر استفاده می‌کند. شرح مفصلی از چارچوب پیشنهادی در بخش زیر ارائه شده است.

3\_5 طبقه بندی MaxNet

مهمترین مرحله در سیستم CBIR پیشنهادی مرحله طبقه بندی است که در آن یک مدل DCNN با استفاده از شیوه های یادگیری نظارت شده برای طبقه بندی تصاویر متعلق به کلاس های مختلف آموزش می بیند. هدف نهایی این مرحله پیش‌بینی برچسب کلاس یک تصویر خاص و همچنین استخراج بردار ویژگی است که در فاز دوم برای بازیابی تصاویر مشابه بصری استفاده می‌شود. مدل MaxNet پیشنهادی در مجموع شامل بیست و یک لایه پیچشی است که به روشی ساختاریافته برای استخراج حداکثر اطلاعات از تصاویر تکرار می‌شوند. یک نمایش تصویری از معماری پیشنهادی در شکل 3\_3 ارائه شده است

 

شکل 3\_3 معماری یادگیری عمیق پیشنهادی MaxNet

 شبکه پیشنهادی، ورودی 4 بعدی دارد که در آن سه بعد صفحه قرمز، سبز و آبی است، در حالی که بعد چهارم، نوع مقیاس خاکستری تصویر است. یک ساختار سه شاخه استفاده می شود که از ظرفیت پردازش موازی سیستم استفاده می کند و سپس خروجی مربوطه را خلاصه می کند. این انباشته شدن موازی لایه ها، ماژول آغازین نامیده می شود و در شکل 4\_3 نشان داده شده است. ماژول آغازین در اصل برای چالش طبقه بندی ILSVRC 2014 معرفی شده است. این شامل 8 لایه کانولوشن است و ویژگی ها را برای ادغام خطوط لوله به هم متصل می کند. ماژول آغازین به روز شده در مقابل شامل سه لایه کانولوشن با هسته است، یعنی 5 × 5،

 3 × 3 و 1 × 1. نقشه‌های ویژگی این هسته‌ها در هر ماژول اولیه متفاوت است تا تنوع در خروجی ویژگی به حداکثر برسد. ماژول اولیه شش بار تکرار می شود تا یک معماری عمیق CNN ایجاد شود، و یک لایه پیچشی حاوی 7 × 7 هسته بعد از هر سه گانه ماژول های آغازین گنجانده شده است. هسته‌های بزرگ اطلاعات متنی بیشتری را اضافه می‌کنند و کیفیت بردار ویژگی را به ماژول اولیه افزایش می‌دهند. بنابراین، این هسته بزرگ تفاوت دیگری را بین ماژول اولیه و ماژول به روز شده پیشنهادی نشان می دهد. تمام لایه های پیچیدگی توسط یک لایه عادی سازی دسته ای دنبال می شوند. لایه عادی سازی دسته ای برای عادی سازی فعال سازی لایه های قبلی استفاده می شود. میانگین مقدار فعال سازی و انحراف استاندارد را به ترتیب نزدیک به صفر و یک حفظ می کند. یک مدل MaxNet عمیق‌تر شامل نه ماژول اولیه نیز امتحان شده است، اما نتایج کمتری را در مقایسه با مدل MaxNet پیشنهادی ایجاد می‌کند. بنابراین، مشخص شده است که افزایش عمق CNN نتایج بهتر را در هر مورد یکسان نمی کند. در بین ماژول های آغازین، یک لایه max-pooling اضافه می شود تا با انتخاب حداکثر مقدار در مجاورت، ویژگی ها را به صورت مستطیلی خلاصه کند. ویژگی های عمیق استخراج شده از ماژول آغازین پس از هر تکرار جمع می شوند. تجمع افزودنی پیشنهاد شده است، زیرا بهترین نتایج تجربی را از نظر اقدامات ارزیابی به همراه دارد. تجمع افزودنی با علامت مثبت در شکل 3\_4 نشان داده شده است.



شکل 3\_4 ماژول اولیه پیشنهادی برای مدل MaxNet

 ذکر شده است که جمع کردن ویژگی‌های عمیق در مقایسه با تفریق، ضرب، بیشینه‌سازی، کمینه‌سازی و الحاق مکمل آن‌ها است.

 چهار نوع گام جمع آوری در بخش های فرعی زیر مورد بحث قرار می گیرد

 3\_5\_1 بیشینه سازی

در این نوع تجمع، یک تابع حداکثر در انتهای ماژول آغازین برای انتخاب ویژگی‌های عمیقی که حداکثر مقادیر را بین سه خط لوله اولیه دارند، استفاده می‌شود. به این ترتیب ویژگی‌های با مقادیر کم یا ناچیز کنار گذاشته می‌شوند و پیچیدگی‌هایی را که در ماژول اولیه یافت می‌شود کاهش می‌دهند. لایه حداکثر سازی نقشه های ویژگی را به عنوان محاسبه می کند:

$O\_{f}$= max(I$P\_{1}$, I$P\_{2}$, I$P\_{3}$)

که در آن، IP1، IP2، و IP3 نشان دهنده نقشه ویژگی عمیق خروجی ماژول اولیه، جریان ورودی یک، جریان ورودی دو و جریان ورودی سه است.

 3\_5\_2 به حداقل رساندن

برخلاف حداکثرسازی، یک تابع حداقل بعد از هر ماژول اولیه برای انتخاب ویژگی های عمیق با حداقل مقادیر استفاده می شود. این نوع تجمع بر روی آزمایش شده است تا نشان دهد که مقادیر کم ویژگی نتایج قابل توجهی برای کار بازیابی تصویر ایجاد نمی کند.

نقشه ویژگی عمیق به حداقل رسیده است به عنوان خروجی:

$O\_{f}$= min(I$P\_{1}$, I$P\_{2}$, I$P\_{3}$)

3\_5\_3 الحاق

الحاق در ماژول اولیه اولیه به عنوان پیشنهاد شده است خوب. در این مرحله، ویژگی‌های عمیق استخراج‌شده از خطوط لوله اولیه به یکدیگر متصل می‌شوند و آرایه‌ای از ویژگی‌ها به اندازه سه آرایه استخراج‌شده از خطوط لوله اولیه ایجاد می‌کنند. به این ترتیب هیچ ویژگی مانند حالت حداکثر سازی و کوچک سازی کنار گذاشته نمی شود. الحاق نتایج خوبی ایجاد می کند اما به قیمت پیچیدگی و کارایی سیستم. این به دلیل این واقعیت است که الحاق ویژگی های مهم و همچنین ناچیز را حفظ می کند و بنابراین دارای تعداد زیادی پارامتر است. بنابراین، خطر تطبیق بیش از حد داده ها را نیز در بر دارد.

نقشه ویژگی عمیق به هم پیوسته بزرگ به عنوان تولید می شود:

$O\_{f}$= concat(I$P\_{1}$, I$P\_{2}$, I$P\_{3}$)

3\_5\_4 تجمع افزودنی

 تجمع افزودنی بهترین نتایج را بدون به خطر انداختن کارایی یا سادگی مدل ایجاد کرد. این مرحله ویژگی های عمیق استخراج شده از خطوط لوله مختلف را اضافه می کند و مقادیر ویژگی را به طور قابل توجهی افزایش می دهد. این عمل منجر به افزایش مقادیر قابل توجه ویژگی می شود و بنابراین قوانین بهتری را برای نقشه برداری مدل گزارش می دهد. از آنجایی که ویژگی ها اضافه شده است، در مقایسه با مرحله الحاق دارای تعداد قابل توجهی کمتری از ویژگی ها است و بنابراین، هم کارآمد و هم ساده است.

 تجمیع افزودنی نقشه های ویژگی عمیق به صورت محاسبه می شود:

$O\_{f}$= add(I$P\_{1}$, I$P\_{2}$, I$P\_{3}$)

خروجی آخرین ماژول آغازین به یک لایه کانولوشن حاوی 5 × 5 هسته و به دنبال آن حداکثر استخر 4 × 4 به نمونه پایین وارد می شود. بردار ویژگی برای تولید احتمالات توزیع کلاس با استفاده از تابع فعال سازی softmax، بیشتر به لایه خروجی متصل است. این احتمالات سافت مکس به عنوان بردارهای ویژگی برای تشکیل یک پایگاه داده ویژگی استفاده می شود. این پایگاه داده در مرحله بازیابی برای تطبیق پرس و جو و بازیابی تصویر استفاده می شود. ویژگی های غیر خطی Softmax SF برای هر برچسب کلاس l به عنوان محاسبه می شود:

SF(c=1|F)=$\frac{e^{FW\_{j}}}{\sum\_{t=1}^{c}e^{XW\_{j}}}$

که در آن F، W، و C بردار ویژگی، پارامترهای وزن و تعداد کل کلاس ها را نشان می دهد، در حالی که، c احتمال کلاس یک تصویر است.

3\_6 بازیابی تصویر

پس از آموزش مدل MaxNet، مجموعه داده وارد شبکه می شود و بردارهای ویژگی از آخرین لایه مدل MaxNet پیشنهادی استخراج می شوند. این بردارهای ویژگی همه تصاویر موجود در مجموعه داده را نشان می دهند. این ویژگی ها همانطور که در شکل 3\_2 نشان داده شده است در پایگاه داده ویژگی ها ذخیره می شوند. تصویر پرس و جو نیز از طریق شبکه پیشخور می شود و ویژگی های استخراج شده با بردارهای ویژگی کل مجموعه داده مقایسه می شود. معیار تشابه مورد استفاده برای مقایسه ویژگی‌ها بین تصویر پرس و جو و بقیه مجموعه داده، اندازه‌گیری فاصله اقلیدسی است. علاوه بر مقایسه ویژگی‌ها، از برچسب کلاس پیش‌بینی‌شده برای محدود کردن ناحیه مورد نظر استفاده می‌شود، بنابراین تصاویر نامربوط نادیده گرفته می‌شوند. همچنین یک راه مفید برای کاهش زمان محاسباتی برای سیستم های بازیابی است. تصاویری که از نظر شاخص شباهت به تصویر پرس و جو نزدیکتر هستند به عنوان نتایج بازیابی شده نمایش داده می شوند.