Key.Net: Keypoint Detection by Handcrafted and Learned CNN Filters

Axel Barroso-Laguna

Edgar Riba

Daniel Ponsa

Krystian Mikolajczyk

{axel.barroso17, k.mikolajczyk}@imperial.ac.uk

{eriba, daniel}@cvc.uab.es

Key.Net: تشخیص Keypoint توسط فیلترهای CNN ساخته شده و آموزش داده شده

**چکیده**

ما یک روش جدید برای تشخیص کلیدواژه معرفی می کنیم. این کار با ترکیبی از فیلترهای CNN ساخته شده و آموخته شده است.

در یک معماری چند مقیاس کم عمق. فیلترهای دست ساز ، ساختارهای لنگر را برای فیلترهای آموخته شده فراهم می کنند که ویژگی های قابل تکرار را محلی، امتیازدهی و رتبه بندی می کنند.

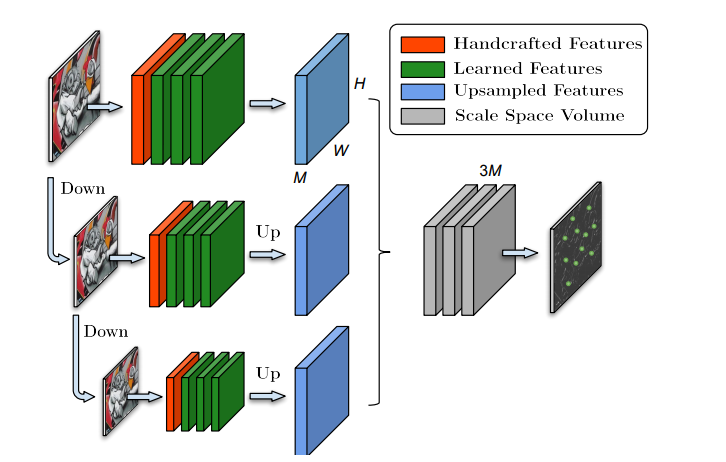
نمایش فضای مقیاس در شبکه برای استخراج نقاط کلیدی یا keypoint در سطوح مختلف استفاده می شود. ما یک تابع از ضرر یا جریمه را برای شناسایی ویژگی های قوی که در طیف وسیعی از مقیاس ها وجود دارد و به حداکثر رساندن نمره تکرارپذیری طراحی می کنیم. مدل Key.Net ما بر روی داده هایی که به صورت مصنوعی از ImageNet ایجاد شده و در معیار HPatches ارزیابی می شود ، آموزش دیده است. نتایج نشان می دهد که روش ما از نظر تکرارپذیری ، عملکرد مطابق و پیچیدگی از advanced detector های فوق العاده عمل می کند.

**مقدمه**

پیشرفت های تحقیقاتی در آشکارسازها و توصیفگرهای ویژگی محلی منجر به پیشرفت های چشمگیری در زمینه هایی مانند تطبیق تصویر ، تشخیص شی ، self-guided navigation یا بازسازی سه بعدی شد. اگرچه جهت کلی روش های تطبیق تصویر در حال حرکت به سمت سیستم های مبتنی بر یادگیری است، اما مزیت روش های یادگیری نسبت به روش های دستی و مرسوم است در تشخیص کلیدواژه نشان داده نشده است.

به طور خاص ، شبکه های عصبی Convolutional (CNN) با وجود ناکارآمدی غیر عملی تکنیک های اولیه ، توانستند به طور قابل توجهی خطای تطبیق را در توصیف کنندگان محلی کاهش دهند. این کارها باعث تحریک تلاش های تحقیقاتی بیشتر و در نتیجه افزایش کارایی توصیف کنندگان مبتنی بر CNN می شود ، برعکس روند کلی علاوه بر موفقیت محدود ردیاب های یاد گرفته شده است.

به طور سنتی ، آشکارسازهای ویژگی محلی بر اساس فیلترهای مهندسی شده ساخته می شدند. به عنوان مثال ، رویکردهایی مانند تفاوت Gaussians ، Harris-Laplace یا Hessian-Affine از ترکیب مشتقات تصویر برای محاسبه نقشه های ویژگی استفاده می کنند ، که به طرز چشمگیری مشابه عملیات در لایه های آموزش دیده CNN است.



*تصویر 1 : معماری پیشنهادی Key.Net ترکیبی از فیلترهای ساخته شده و آموخته شده برای استخراج ویژگی ها در مقیاس های مختلف است. نقشه های ویژگی نمونه برداری و منشور می شوند. فیلتر آخرین آموخته شده برای به دست آوردن final response map ، میزان فضای مقیاس را ترکیب می کند.*

فقط با چند لایه ، یک شبکه می تواند با یادگیری مقادیر مناسب در فیلترهای کانولوشن ، رفتار ردیاب های سنتی را تقلید کند. با این حال ، بر خلاف موفقیت در توصیفگرهای تصویر محلی مبتنی بر CNN ، پیشرفت در آشکارسازهای دست ساز ارائه شده توسط روش های کاملاً مبتنی بر CNN از نظر معیارهای پذیرفته شده گسترده ای مانند تکرارپذیری محدود است. یکی از دلایل دقت کم آنها هنگام تخمین پارامترهای ترکیبی مناطق مشخصه است. مقاومت در برابر تغییرات مقیاس به ویژه مشکل ساز به نظر می رسد در حالی که پارامترهای دیگر مانند جهت گیری غالب را می توان به خوبی توسط CNNs پس گرفت این باعث ایجاد انگیزه در معماری جدید ما می شود که Key.Net نامیده می شود و از فیلترهای دست ساز و آموخته شده و همچنین نمایش چند مقیاس استفاده می کند. معماری Key.Net در شکل 1 نشان داده شده است. معرفی فیلترهای دست ساز ، که به عنوان لنگرهای نرم عمل می کنند ، امکان کاهش تعداد پارامترهای استفاده شده توسط ردیاب های پیشرفته را فراهم می کند ، در حالی که از نظر تکرارپذیری عملکرد را حفظ می کند. این مدل با بازنمایی چند مقیاس از تصاویر در اندازه کامل عمل می کند و یک نقشه پاسخ شامل نمره نقطه کلیدی برای هر پیکسل برمی گرداند. ورودی چند مقیاس به شبکه اجازه می دهد تا نقاط کلیدی پایدار را در مقیاس ها ارائه دهد ، بنابراین قدرت را برای تغییرات مقیاس فراهم می کند.

در حالت ایده آل ، یک robust detector قادر است ویژگی های یکسانی را برای تصاویری که تحت تغییر شکل هندسی یا فوتومتریک قرار دارند ، ارائه دهد. تعدادی از کارهای مرتبط عملکرد هدف خود را برای پرداختن به این موضوع متمرکز کرده اند ، اگرچه این کارها یا براساس تکه های محلی [9 ، 10] یا از دست دادن رگرسیون نقشه جهانی بود. در مقابل ، ما از دست دادن محدودیت متغیر را به یک عملکرد هدف جدید گسترش می دهیم که اطلاعات محلی و جهانی را با هم ترکیب می کند. ما یک اپراتور کاملاً متفاوت ، پیشنهاد شاخص چند مقیاس را طراحی می کنیم که نقاط کلیدی را در مناطق چند مقیاس پیشنهاد می کند. ما به طور گسترده روش در معیار HPatches اخیرا را از نظر دقت و تکرار مطابق با پروتکل از ارزیابی می کنیم.

به طور خلاصه ، مشارکت های ما به شرح زیر است:

الف) یک keypoint detector ترکیبی از ویژگی های CNN آموزش دیده و ویژگی های دستی است.

ب) یک ضرر و عملگر جدید در مقیاس چند منظوره برای شناسایی و رتبه بندی کلیدهای پایدار در مقیاس ها

ج) تشخیص ویژگی چند مقیاس با معماری کم عمق.

بقیه مقاله به شرح زیر است. ما کار مربوطه را در بخش 2 بررسی می کنیم. بخش 3 معماری ترکیبی پیشنهادی Key.Net فیلترهای CNN ساخته شده و آموخته شده را ارائه می دهد و بخش 4 ضرر را معرفی می کند. جزئیات اجرا و آزمایشی در بخش 5 آورده شده و نتایج در بخش 6 ارائه شده است.

**کارهای مشابه**

نظرسنجی های بسیاری وجود دارد که به طور گسترده در مورد روش های تشخیص ویژگی بحث می کند. ما کارهای مرتبط را در دو دسته اصلی ارائه می دهیم:

* handcrafted based.
* learned based