شبکه‌های کانولوشنی مستقل از مختصات

کانولوشن‌های یکسان نسبت به ایزومتری و پیمانه روی منیفلدهای ریمانی

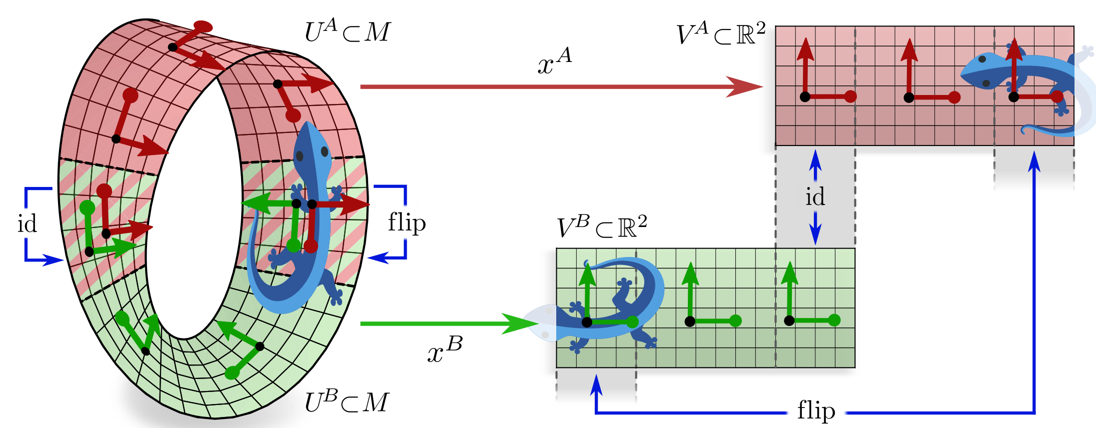
موریس وایلر[۱] پاتریک فوره[۱] اریک ورلینده[۱] مکس ولینگ[۱,۲]

[۱]دانشگاه آمستردام
[۲]پژوهشگاه هوش مصنوعی کوالکام

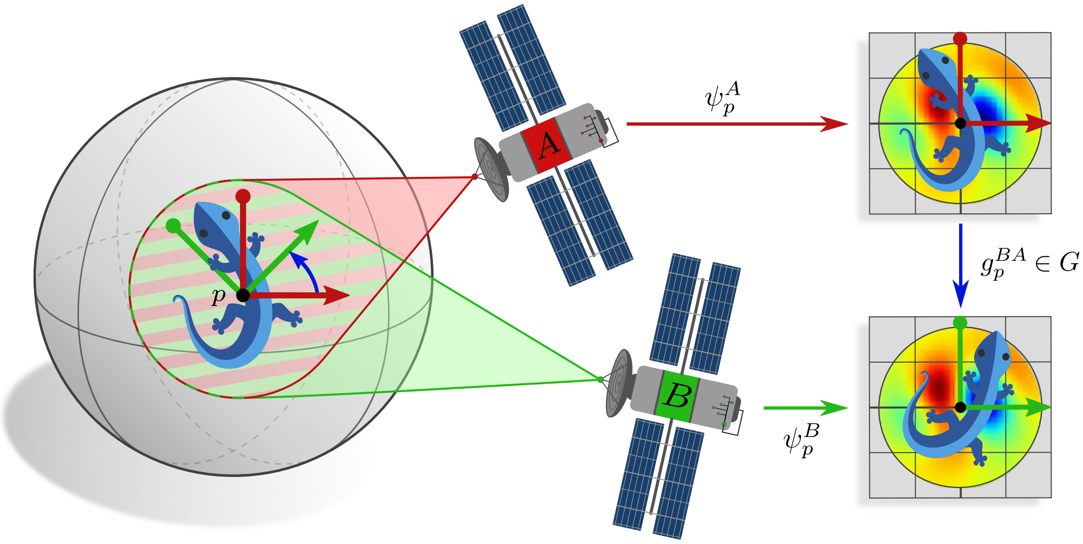
p.d.forre@uva.nl m.weiler.ml@gmail.com
m.welling@uva.nl e.p.verlinde@uva.nl

□□□□□□□

انگیزه‌ای از موفقیت گسترده شبکه‌های عمیق کانولوشنی، علاقه زیادی برای تعمیم کانولوشن‌ها به منیفلدهای غیراقلیدسی وجود دارد. یک پیچیدگی عمده در مقایسه با فضاهای مسطح این است که مشخص نیست کرنل کانولوشن باید در کدام تراز روی یک منیفلد اعمال شود. دلیل اساسی این ابهام آن است که منیفلدهای عمومی دارای انتخاب متعارف چارچوب‌های مرجع (گیج) نیستند. بنابراین کرنل‌ها و ویژگی‌ها باید نسبت به مختصات دلخواه بیان شوند. ما استدلال می‌کنیم که انتخاب خاص مختصات‌بندی نباید بر استنتاج شبکه تأثیر بگذارد □ آن باید مستقل از مختصات باشد. تقاضای همزمان برای استقلال مختصات و اشتراک وزن منجر به الزامی روی شبکه می‌شود تا تحت تبدیل‌های گیج محلی (تغییرات چارچوب‌های مرجع محلی) تناوب‌پذیر باشد. ابهام چارچوب‌های مرجع بدین‌گونه به G-ساختار منیفلود بستگی دارد، به طوری که سطح لازم تناوب‌پذیری گیج توسط گروه ساختار G متناظر تجویز می‌شود. کانولوشن‌های مستقل از مختصات ثابت می‌شوند که نسبت به آن ایزومتری‌هایی که تقارن‌های G-ساختار هستند تناوب‌پذیر باشند. نظریه حاصل به شکل آزاد از مختصات بر حسب بندل‌های فیبر فرمول‌بندی می‌شود. برای نمونه‌سازی طراحی کانولوشن‌های مستقل از مختصات، ما شبکه کانولوشنی روی نوار موبیوس پیاده‌سازی می‌کنیم. عمومیت فرمول‌بندی هندسه دیفرانسیل ما از شبکه‌های کانولوشنی با بررسی گسترده ادبیات نشان داده می‌شود که تعداد زیادی از CNNهای اقلیدسی، CNNهای کروی و CNNها روی سطوح عمومی را به عنوان نمونه‌های خاص کانولوشن‌های مستقل از مختصات توضیح می‌دهد.



شکل ۱: نمایش کانولوشن روی نوار موبیوس با گیج‌های مختلف



شکل ۲: مشاهده‌گران مختلف A و B ممکن است یک الگوی از ویژگی‌ها را از «دیدگاه» متفاوتی درک کنند. ماهواره‌ها در کاربرد ما کرنل‌های کانولوشنی هستند که میدان دید محلی خود را در اطراف p به یک بردار ویژگی در p خلاصه می‌کنند. «دیدگاه» آنها انتخاب یک چارچوب مرجع محلی (گیج) در p است که کرنل بر اساس آن تراز می‌شود. از آنجا که مشاهدات هر دو دیدگاه نشان‌دهنده یک الگوی یکسان هستند، پاسخ‌های کرنل باید حاوی اطلاعات معادل باشند، یعنی استنتاج باید مستقل از مختصات باشد. این امر کرنل‌های کانولوشنی را ملزم می‌کند تا تحت تبدیل‌های گیج محلی (یعنی تغییرات چارچوب‌های مرجع) تناوب‌پذیر باشند. سطح تناوب‌پذیری گیج توسط گروه ساختار G تعیین می‌شود، که هم به منیفولد و هم به کاربرد بستگی دارد.

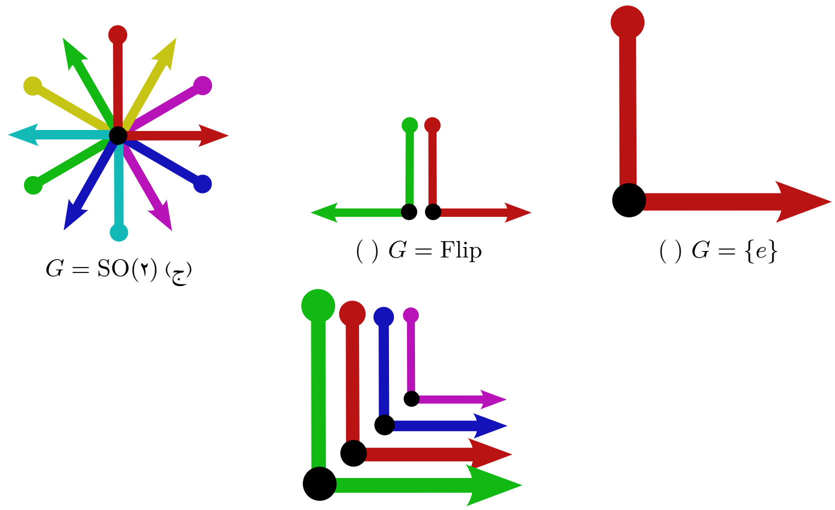
۱ مقدمه

در سال‌های اخیر، شبکه‌های عصبی عمیق به مدل‌های منتخب برای طیف گسترده‌ای از وظایف در یادگیری ماشین تبدیل شده‌اند. موفقیت مدل‌های عمیق اغلب ریشه در طراحی وظیفه–خاص دارد که منعکس‌کننده ساختار ریاضی داده‌های مورد پردازش است. یک مثال برجسته شبکه‌های کانولوشنی () هستند که از طریق اتصال محلی و اشتراک وزن فضایی، ساختار مکانی داده‌ها را بهره‌برداری می‌کنند. از آنجا که یک کرنل یکسان در هر نقطه از فضا اعمال می‌شود، شبکه‌های کانولوشنی نسبت به انتقال تناوب‌پذیر هستند، به این معنی که الگوهای آموخته‌شده را به طور خودکار بر روی موقعیت‌های فضایی تعمیم می‌دهند. با توجه به موفقیت تجربی قابل توجه اقلیدسی، علاقه زیادی به تعمیم مدل‌های کانولوشنی برای پردازش سیگنال‌ها در دامنه‌های عمومی‌تر و تناوب‌پذیر ساختن آنها تحت گروه‌های تقارن بزرگ‌تر وجود دارد.

این کار به بررسی تعمیم شبکه‌های کانولوشنی به منیفلدهای ریمانی می‌پردازد. یک پیچیدگی عمده در تعمیم شبکه‌های کانولوشنی از فضاهای اقلیدسی ^{d} به منیفلدهای ریمانی عمومی این است که منیفلدهایی دارای انتخاب مرجع جهت ارجح نیستند که کرنل کانولوشن بتواند بر اساس آن برای اندازه‌گیری ویژگی‌ها تراز شود. از آنجا که هیچ جهت مرجعی ترجیح داده نمی‌شود، کرنل باید به صورت دلخواه بر روی منیفولد تراز شود. موضوع اصلی این کار تنظیم این اختیار با مستقل ساختن استنتاج شبکه‌ها از تراز خاص کرنل‌های کانولوشن است. مشخص می‌شود که این امر مستلزم آن است که کرنل‌ها تناوب‌پذیر گیج باشند، یعنی تحت تبدیل‌های تراز کرنل، تناوب‌پذیر باشند. پاسخ یک کرنل تناوب‌پذیر گیج، هنگام تغییر تراز آن، به طور قابل پیش‌بینی تغییر می‌کند – بنابراین محتوای اطلاعات استخراج شده برای هر انتخاب (دلخواه) تراز، یکسان تضمین می‌شود.

ما تراز یک کرنل در نقطه‌ای p از یک منیفولد M را به عنوان انتخاب یک چارچوب مرجع محلی □ یا گیج □ از فضای مماس مربوطه T_pM رسمی‌سازی می‌کنیم. بنابراین تبدیل‌های گیج، تبدیلاتی بین انتخاب‌های چارچوب‌های مرجع هستند. شکل ؟؟ مفهوم تراز کردن کرنل‌ها در امتداد چارچوب‌های مرجع را بصری‌سازی می‌کند. تراز کردن کرنل نسبت به میدان چارچوب کانونی (منحصر به فرد ارجح) صفحه اقلیدسی [۲]، همانطور که در بالا نشان داده شده است، منجر به میدان کرنل معمول اقلیدسی می‌شود. یک میدان چارچوب متفاوت، همانطور که در پایین نشان داده شده است، یک میدان کرنل و در نتیجه یک شبکه جایگزین را نشان می‌دهد. همانطور که در بالا ذکر شد، در اکثر منیفلدهایی، انتخاب چارچوب‌ها ذاتاً مبهم است به طوری که هیچ تراز کرنل خاصی ترجیح داده نمی‌شود. شکل ۲ این مشکل را برای کره S^2 بصری‌سازی می‌کند، جایی که چارچوب‌ها فقط تا چرخش‌ها منحصر به فرد هستند.

سطح ابهام در انتخاب چارچوب‌های مرجع به ساختار هندسی منیفولد بستگی دارد. چنین ساختاری اغلب امکان رفع ابهام چارچوب‌های مرجع تا تبدیل‌های تقارنی خاص (تبدیل‌های گیج) را فراهم می‌کند؛ به شکل ۳ مراجعه کنید. این گزاره با چند مثال بهتر توضیح داده می‌شود:



(ج) $G = \mathrm{SO}(۲)$ ( ) $G = \mathrm{Flip}$ ( ) $G = \{e\}$

( ) $G = \mathrm{Scale}$

شکل ۳: انتخاب چارچوب‌های مرجع یک فضای مماس T_pM همیشه منحصر به فرد نیست. ساختار هندسی (G-ساختار) یک منیفولد، زیرمجموعه‌ای ترجیحی از چارچوب‌های مرجع را نشان می‌دهد به طوری که تبدیل‌های گیج بین این چارچوب‌ها در گروه ساختار $G \leq \mathrm{GL}(d)$ قرار می‌گیرند. اشکال ??، ??، ۳ج و ?? چنین زیرمجموعه‌هایی از چارچوب‌ها را به ترتیب برای گروه بدیهی $G = \{e\}$، گروه بازتاب $G = $ □□□□، گروه چرخش $G = \mathrm{SO}(۲)$ و گروه مقیاس $G = $ □□□□□ نشان می‌دهند. ویژگی‌ها، اندازه‌گیری‌ها را نسبت به هر یک از چارچوب‌های متمایز کدگذاری می‌کنند. ضرایب عددی آن‌ها نسبت به چارچوب‌های مختلف با عمل یک نمایش گروهی ρ از G مرتبط هستند.

- یک منیفولد هموار هیچ اولویتی در انتخاب چارچوب‌ها ندارد. تبدیل‌های گیج بین چارچوب‌های عمومی، نگاشت‌های خطی معکوس‌پذیر دلخواه هستند، یعنی مقادیری در گروه خطی عمومی $G = \mathrm{GL}(d)$ می‌گیرند.
- یک جهت‌گیری منیفولد امکان تشخیص چارچوب‌های راست‌گرد از چپ‌گرد را می‌دهد. تبدیل‌های گیج بین چارچوب‌های هر دو دست، جهت‌گیری را حفظ می‌کنند، یعنی اعضایی از $G = \mathrm{GL}^+(d)$ هستند (نگاشت‌های خطی معکوس‌پذیر با دترمینان مثبت).
- یک فرم حجم امکان تشخیص چارچوب‌های واحد حجم را می‌دهد. تبدیل‌های گیج سپس حجم را حفظ می‌کنند، یعنی مقادیری در گروه خطی خاص $G = \mathrm{SL}(d)$ می‌گیرند.
- ساختار متریک یک منیفولد ریمانی امکان اندازه‌گیری فواصل و زوایا در فضاهای مماس را فراهم می‌کند و بنابراین امکان تشخیص چارچوب‌های متعامد را می‌دهد. تبدیل‌های گیج بین چارچوب‌های متعامد، چرخش‌ها و بازتاب‌ها در گروه متعامد $G = \mathrm{O}(d)$ هستند.
- همچنین، یک جهت‌گیری و متریک به معنای چارچوب‌های متعامد جهت‌دار است. تبدیل‌های گیج سپس فقط چرخش‌ها در گروه متعامد خاص $G = \mathrm{SO}(d)$ هستند.
- یک میدان چارچوب بر روی منیفولد شامل یک چارچوب منحصر به فرد در هر نقطه از منیفولد است. تبدیل‌های گیج در این حالت بدیهی هستند، که توسط گروه بدیهی $G = \{e\}$ توصیف می‌شود.

همه این ساختارهای هندسی یک ویژگی مشترک دارند که یک زیرمجموعه (زیربندل) ترجیحی از چارچوب‌ها را تعریف می‌کنند به طوری که تبدیل‌های گیج مقادیری در یک گروه ساختار $G \leq \mathrm{GL}(d)$ می‌گیرند. برای تأکید بر نقش مرکزی گروه ساختار G، چنین ساختارهایی به عنوان G–ساختارها $\mathcal{M}$ نامیده می‌شوند. مثال‌های بصری از G–ساختارها برای گروه‌های ساختار G و منیفلدهای M مختلف در شکل ?? ارائه شده‌اند.

از آنجا که انتخاب چارچوب‌های مرجع ذاتاً مبهم است، هر کمیت هندسی و عملیات شبکه باید به همان اندازه به خوبی نسبت به چارچوب‌های دلخواه G–ساختار $\mathcal{M}$ قابل نمایش باشند، یعنی باید مستقل از مختصات $\mathcal{M}$ باشند. بنابراین بردارهای ویژگی با یک نمایش گروهی (عمل گروهی خطی) ρ از گروه ساختار G مرتبط هستند که قانون تبدیل آن‌ها را تحت تبدیل‌های گیج (انتقال‌های با مقدار G بین چارچوب‌های مرجع) تعیین می‌کند. انتخاب خاص نمایش گروهی، نوع هندسی میدان بردار ویژگی را تعیین می‌کند. مثال‌های معمول شامل میدان‌های اسکالر، بردار یا تانسور هستند، با این حال، انواع میدان‌های عمومی‌تر نیز در عمل استفاده می‌شوند. شکل ?? استقلال مختصات کمیت‌های هندسی را در مثال شناخته شده بردارهای مماس بصری‌سازی می‌کند.

هر لایه شبکه باید قوانین تبدیل ویژگی‌ها را رعایت کند، یعنی باید تضمین کند که خروجی‌های آن طبق انتظار تبدیل می‌شوند. به طور خاص برای کانولوشن‌ها، استقلال مختصات $\mathcal{M}$ ایجاب می‌کند که اعمال کرنل مشترک نسبت به چارچوب‌های مختلف G–ساختار در نقطه‌ای $p \in M$ باید همان