

## سوال 2 عملی قسمت تئوری :

( الف )

شبکه‌های کانولوشنی عادی از یک sampling grid (ترکیب موقعیت‌های مختلف در تصویر) ثابت برای اعمال فیلترهای کانولوشنی استفاده می‌کنند. این به این معنی است که هر عنصر در لایه‌ی خروجی با یک مکان مشخص در لایه‌ی ورودی متناظر است.

در حالی که در شبکه‌های کانولوشنی Deformable، امکان استفاده از یک sampling grid تطبیق‌پذیر و قابل یادگیری فراهم می‌شود. این به شبکه این امکان را می‌دهد که برای هر ناحیه از تصویر، یک sampling grid مناسب و بهینه را انتخاب کند. این قابلیت به شبکه این اجازه را می‌دهد که بهترین موقعیت‌ها را برای اعمال فیلترها یا عملیات کانولوشنی بیابد.

در نتیجه، شبکه‌های کانولوشنی Deformable می‌توانند در مواردی که ویژگی‌ها یا الگوهای مهم در تصویر در مکان‌های مختلف و با اندازه‌های مختلف قرار دارند، عملکرد بهتری داشته باشند. این خصوصیت می‌تواند در وظایفی مانند تشخیص اشیاء با اندازه‌ها و مکان‌های مختلف یا تشخیص لبه‌ها و یا ویژگی‌های هندسی پیچیده مفید باشد.

( ب )

شبکه‌های کانولوشنی Deformable از لایه‌هایی به نام "Deformable Convolution" برای انجام تغییرات هندسی انعطاف‌پذیر در تصاویر استفاده می‌کنند. این لایه‌ها امکان اعمال تغییرات جابجایی (offsets) به فیلترها (کرنل‌ها) را می‌دهند تا شبکه بتواند بهتر به الگوها و ویژگی‌های موجود در تصویر واکنش نشان دهد. در ادامه توضیحاتی در مورد چگونگی انعطاف‌پذیری در شبکه‌های کانولوشنی Deformable آورده شده است:

## 1. Sampling Grids و Offsets

- ابتدا، برای هر نقطه در ورودی، یک offset مربوط به این نقطه محاسبه می‌شود. این offset نشان‌دهنده جابجایی یا تغییر مکان مورد نظر است.

- با استفاده از این offsetها، یک sampling grid ایجاد می‌شود. این sampling grid نشان‌دهنده موقعیت‌های جدیدی است که فیلترها در آن‌ها باید بر روی تصویر اعمال شوند.

## 2. Deformable Convolution

- با استفاده از این sampling grid، اعمال کانولوشن به صورت معمول صورت می‌گیرد. با این تفاوت که هر فیلتر به صورت پویا و با توجه به offsetها و sampling grid جابجا می‌شود.

- این انعطاف‌پذیری به شبکه این امکان را می‌دهد که الگوها را در مکان‌ها و اندازه‌های مختلف تشخیص دهد.

## 3. آموزش Offsetها

- Offsetها به صورت قابل یادگیری آموزش داده می‌شوند. این به مدل این امکان را می‌دهد که در طول فرآیند آموزش، بهینه‌ترین offsetها را برای تصاویر خاصی کشف کرده و استفاده کند.

به این ترتیب، شبکه کانولوشنی Deformable از این تکنیک‌ها استفاده می‌کند تا با تغییرات هندسی انعطاف‌پذیر در تصاویر مقابله کرده و ویژگی‌های مفیدتری را در تصاویر استخراج کند. این امکان به شبکه کمک می‌کند تا به مکان‌ها و ویژگی‌های مختلف تصاویر واکنش نشان دهد و بهبود در وظایفی مانند تشخیص اشیاء با مکان‌ها و اندازه‌های مختلف را فراهم کند.

(ج)

شبکه‌های کانولوشنی ساده در مواجهه با تصاویری که شامل تغییرات فضایی زیادی مثل چرخش و انعکاس هستند، ممکن است با مشکلاتی مواجه شوند. دلایل اصلی این مشکلات به شرح زیر هستند:

## 1. Invariance to Transformation

- شبکه‌های کانولوشنی ساده به صورت طبیعی تغییرات جزئی مکانی (translational invariance) را مدل می‌کنند. این به این معنی است که اگر یک شیء در تصویر جابه‌جا شود، تا حدی مدل قابلیت تشخیص آن را دارد. اما برای تغییرات مکانی بزرگتر مثل چرخش یا انعکاس، این توانایی محدود می‌شود.

## 2. تغییرات ژئومتریک

- چرخش و انعکاس تصاویر تغییرات ژئومتریکی هستند که ممکن است باعث تغییر شکل و موقعیت نسبت به محورهای مختصات شود. شبکه‌های کانولوشنی ساده نمی‌توانند به خوبی این تغییرات ژئومتریک را مدل کنند.

## 3. تغییر مکان‌های مهم

- در صورت چرخش یا انعکاس تصویر، مکان‌های مهمی ممکن است به مکان‌های دیگر منتقل شوند. این مسئله باعث می‌شود که فیچرهای مهم در تصویر در مکان‌های نادرستی استخراج شوند و توانایی تشخیص الگوها و ویژگی‌ها کاهش یابد.

## 4. Overfitting to Specific Transformations

- شبکه‌های کانولوشنی ساده ممکن است به خاطر محدودیت در توانایی یادگیری تغییرات ژئومتریک، به صورت زیادی به تغییرات خاص یک دیتاست وابسته شوند و در مقابل تغییرات جدید دچار مشکلات شوند.

برای مقابله با این مشکلات، روش‌هایی همچون استفاده از شبکه‌های کانولوشنی Deformable که امکان انجام تغییرات هندسی انعطاف‌پذیرتر را فراهم می‌کنند، یا استفاده از تکنیک‌هایی مانند داده‌افزایی با چرخش و انعکاس می‌توانند مفید باشند. همچنین، شبکه‌هایی که با استفاده از ترکیبات از قبل آموزش دیده شده (pretrained) بر روی دیتاهای متنوع، اغلب توانمندی بهتری در تشخیص الگوها در تصاویر متغیر دارند.

در شبکه‌های کانولوشنی دیفرمیبل (Deformable Convolutional Networks یا DCN)، از یک لایه خاص به نام "Deformable Convolution" برای اجرای عملیات کانولوشن با تغییرات مکانی انعطاف‌پذیر استفاده می‌شود. یکی از عناصر مهم در این لایه، محاسبه آفست‌ها (offsets) است که نشان‌دهنده جابه‌جایی مکانی هر نقطه نسبت به موقعیت مرجع است.

محاسبه آفست‌ها در لایه Deformable Convolution به صورت یادگیری‌پذیر (learnable) انجام می‌شود. به طور کلی، فرآیند محاسبه آفست‌ها به شکل زیر است:

#### 1. محاسبه مپ ویژگیهای آفست

- ابتدا برای هر نقطه در ورودی، یک مپ ویژگی برای محاسبه آفست‌ها ایجاد می‌شود. این مپ ویژگی‌ها به عنوان ورودی به یک لایه یادگیری آفست می‌رود.

#### 2. یادگیری آفست‌ها

- لایه یادگیری آفست از مپ ویژگیهای ایجاد شده استفاده می‌کند و آفست‌ها را بر اساس این ورودیها یاد می‌گیرد.

- آفست‌ها از نظر تغییرات مکانی مختلف در تصویر یاد گرفته می‌شوند.

#### 3. اعمال آفست به sampling grid

- آفست‌ها به یک sampling grid اعمال می‌شوند. این sampling grid نشان‌دهنده موقعیت‌های جدیدی است که کرنل‌ها (فیلترها) باید بر روی تصویر اعمال شوند.

#### 4. عملیات کانولوشن معمولی

- پس از محاسبه آفست‌ها و اعمال به sampling grid، عملیات کانولوشن به صورت معمول با استفاده از این sampling grid

انجام می‌شود. این باعث می‌شود که کرنل‌ها در موقعیت‌های مختلف اعمال شوند و انعطاف‌پذیری مکانی بهبود یابد.

به عنوان مثال، اگر  $(x, y)$  مختصات یک نقطه در تصویر باشد و  $(\Delta x, \Delta y)$  آفست‌های محاسبه شده برای این

نقطه باشند، موقعیت جدید نقطه محاسبه شده به صورت زیر است:

$$(x_{new}, y_{new}) = (x + \Delta x, y + \Delta y)$$

این امکان به شبکه کمک می‌کند که در تغییرات مکانی مختلف به خوبی عمل کرده و ویژگی‌های مرتبط با انواع مختلف از اشیاء را

در تصویر استخراج کند.