Advanced Neural Networks

Applications-Focused Introduction for Beginners
CMSC 173 - Machine Learning

Noel Jeffrey Pinton
October 16, 2025

Department of Computer Science
University of the Philippines - Cebu

Outline

Introduction: Why Advanced Neural Networks?

Convolutional Neural Networks (CNNs)

Generative Models: Creating New Content

GANs: Generative Adversarial Networks

VAEs: Variational Autoencoders

Transformers: The Revolution in AI

Diffusion Models: The Newest Revolution

Ethics & Responsible AI

Key Concept: Attention Mechanism

Using These Models: Practical Guide

Summary & Looking Forward

Introduction: Why Advanced Neural Networks?

What are Advanced Neural Networks?

Basic Neural Networks

- Fully connected layers
- Good for tabular data
- Limited to simple patterns
- We learned these already!

Advanced Architectures

- **CNNs:** For images and spatial data
- **Transformers:** For text and sequences
- **GANs:** Generate new data
- **VAEs:** Learn compressed representations
- **Diffusion:** Create high-quality images

Why Learn These?

They power the AI you use every day:

- ChatGPT (Transformer)
- DALL-E 2 (Diffusion)
- Face unlock on phones (CNN)
- Google Translate (Transformer)
- AI art generators (GAN/Diffusion)

This Module's Focus

Understanding applications rather than complex math!

CNN Applications: Where You See Them Every Day!

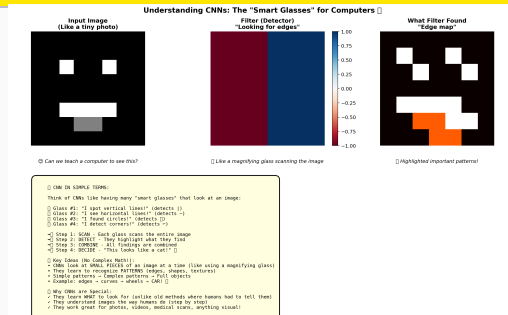|		
Photo Recognition *Your phone:* *'This is a beach!'*	**Self-Driving Cars** *Detecting pedestrians,* *stop signs, other cars*	**Medical Diagnosis** *Finding tumors in* *X-rays and MRI scans*
Face Unlock *Your phone recognizes* *YOUR face to unlock*	**Image Filters** *Snapchat/Instagram* *fun face filters*	**Visual Search** *Google Lens:* *Find similar products*

Course Philosophy

Learn by seeing what's possible! We'll focus on understanding what these networks can do and how to use them, not deriving complex mathematics.

Convolutional Neural Networks (CNNs)

Understanding CNNs: The "Smart Glasses" for Computers

Input Image (Like a tiny photo) — ☺ Can we teach a computer to see this?
Filter (Detector) "Looking for edges" — Like a magnifying glass scanning the image
What Filter Found "Edge map" — Highlighted important patterns!

Simple Explanation

CNNs are neural networks designed for images. They work by:

- Looking at small patches of the image
- Finding patterns (edges, shapes, textures)
- Building up to complex objects
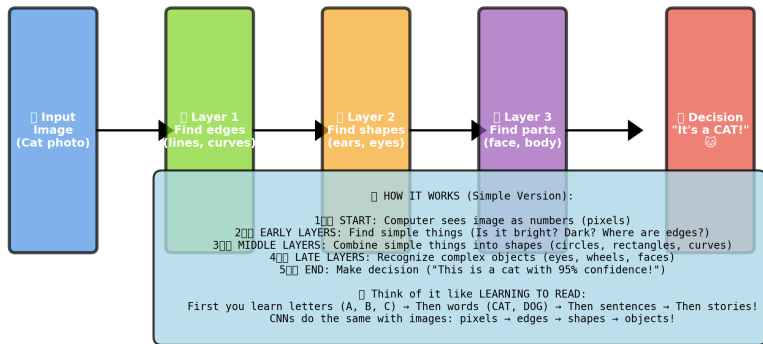- Making decisions based on what they see

Why Not Regular NNs?

- Images have too many pixels

Key Insight

CNNs learn to recognize patterns automatically - no manual feature engineering!

CNN Architecture: From Simple to Complex (Layer by Layer)

🖼 Input Image (Cat photo) → **🔲 Layer 1 Find edges** (lines, curves) → **🟠 Layer 2 Find shapes** (ears, eyes) → **🟣 Layer 3 Find parts** (face, body) → **🎯 Decision "It's a CAT!"** 🐱

💡 HOW IT WORKS (Simple Version):

1️⃣ START: Computer sees image as numbers (pixels)
2️⃣ EARLY LAYERS: Find simple things (Is it bright? Dark? Where are edges?)
3️⃣ MIDDLE LAYERS: Combine simple things into shapes (circles, rectangles, curves)
4️⃣ LATE LAYERS: Recognize complex objects (eyes, wheels, faces)
5️⃣ END: Make decision ("This is a cat with 95% confidence!")

📖 Think of it like LEARNING TO READ:
First you learn letters (A, B, C) → Then words (CAT, DOG) → Then sentences → Then stories!
CNNs do the same with images: pixels → edges → shapes → objects!

Processing Pipeline

Input Image → Find Edges → Find Shapes → Find Objects → Decision

Analogy

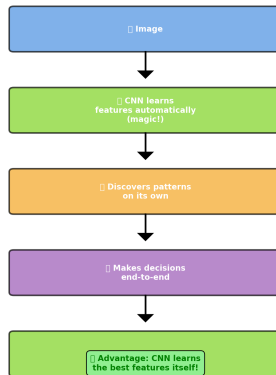Like how humans see: First we see lines and edges, then shapes, then we recognize "this is a cat!"

Why CNNs Changed Everything

OLD WAY (Before CNNs)	NEW WAY (CNNs)
Image	Image
↓	↓
Humans design feature detectors (hard work!)	CNN learns features automatically (magic!)
↓	↓
Extract features manually	Discovers patterns on its own
↓	↓
Simple classifier (like decision tree)	Makes decisions end-to-end
↓	↓
Problem: Humans had to figure out what features to look for!	Advantage: CNN learns the best features itself!
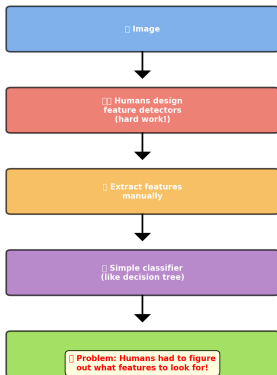
Traditional Methods

- Manual feature design
- Hard to adapt to new tasks
- Limited accuracy
- Lots of expert knowledge needed

CNNs

- Automatic feature learning
- Easily adapt to new problems
- State-of-the-art accuracy
- Just need training data

7

Cancer Detection

Real Application:

- Detect tumors in X-rays and MRIs
- Classify skin lesions (benign/malignant)
- Analyze mammograms for breast cancer
- Help radiologists work faster

Impact

- Earlier disease detection
- Fewer missed diagnoses
- Reduced radiologist workload
- Available in rural areas

Retinal Disease Diagnosis

Example: Google's Diabetic Retinopathy Detection

- Analyzes eye scans
- Detects diabetes complications
- Matches expert doctor accuracy
- Used in India, Thailand

Success Story

FDA-approved AI systems now assist doctors in real hospitals!

CNN Applications: Self-Driving Cars

Lane Detection

What CNNs Do:

- Identify road lane markings
- Track lane boundaries in real-time
- Work in various lighting conditions
- Handle curves and intersections

Object Detection

- Detect pedestrians, cars, cyclists
- Recognize traffic signs and lights
- Estimate distance to objects
- Predict object movement

Companies Using This

- **Tesla:** Full Self-Driving (FSD)
- **Waymo:** Autonomous taxis
- **Cruise:** Robotaxis in SF
- **Mobileye:** Driver assistance

Real Deployment

Over 1 million vehicles use CNN-based vision systems today!

Phone Unlock (Face ID)

How It Works:

- CNN extracts facial features
- Creates unique "face print"
- Compares to stored template
- Works in different lighting
- Adapts to appearance changes

Daily Use Cases

- iPhone/Android face unlock
- Photo organization (Google Photos)
- Security access control
- Airport immigration

Social Media Applications

- **Facebook:** Auto-tag friends in photos
- **Snapchat:** Face filters and effects
- **Instagram:** Beauty filters
- **TikTok:** Face tracking for AR

Privacy Note

Face recognition raises important privacy concerns - always consider ethics!

Smart Security Cameras

Capabilities:

- Detect people vs animals
- Recognize package delivery
- Identify suspicious behavior
- Track movement patterns
- Send targeted alerts

Consumer Products

- Ring Doorbell cameras
- Nest security systems
- Arlo smart cameras
- Reduce false alarms by 90%

Retail Applications

Amazon Go Stores:

- Track what customers pick up
- Automatic checkout (no cashiers)
- Prevent shoplifting
- Analyze shopping behavior

Industry Impact

Checkout-free stores save 75% of labor costs while improving customer experience!

Environmental Monitoring

Applications:

- Track deforestation in Amazon
- Monitor crop health
- Detect illegal fishing
- Assess disaster damage
- Map urban growth

Real Projects

- **Planet Labs:** Daily Earth imaging
- **Global Fishing Watch:** Ocean monitoring
- **NASA:** Climate change tracking

Humanitarian Uses

- Count refugees in camps
- Assess natural disaster impact
- Map poverty indicators
- Monitor conflict zones
- Guide relief efforts

Scale

CNNs can analyze millions of satellite images - impossible for humans alone!
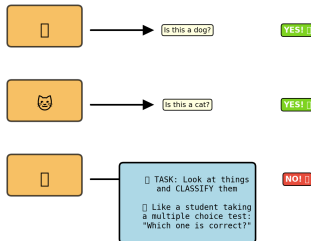
Generative Models: Creating New Content

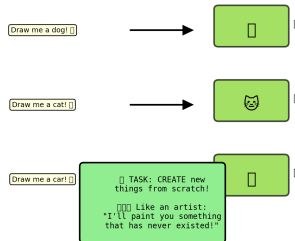Generative vs Discriminative: Creating vs Classifying

DISCRIMINATIVE MODELS

(The Classifier - "What is this?")

GENERATIVE MODELS

(The Creator - "Make me something new!")

Discriminative Models

What they do:

- Classify/label existing data
- "Is this a cat or dog?"
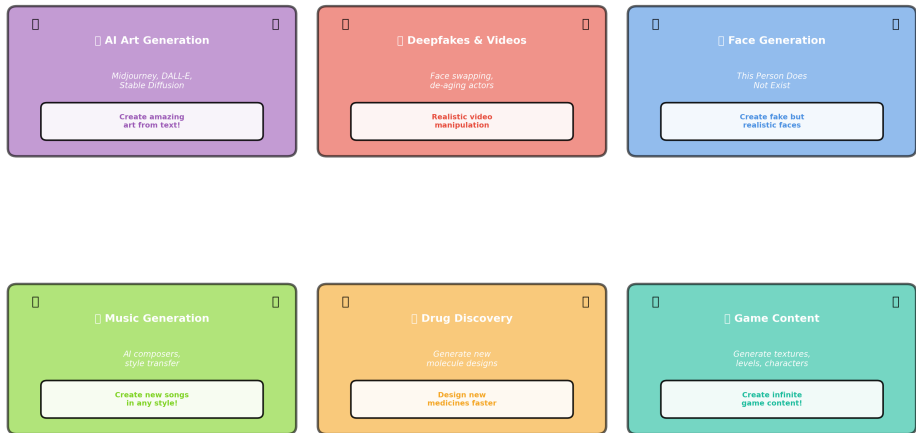- CNNs for image classification

Generative Models

What they do:

- Create new data
- "Generate a new cat image"
- GANs, VAEs, Diffusion models

Generative Models: Creating the Future!

AI Art Generation	**Deepfakes & Videos**	**Face Generation**
Midjourney, DALL-E, Stable Diffusion	*Face swapping, de-aging actors*	*This Person Does Not Exist*
Create amazing art from text!	**Realistic video manipulation**	**Create fake but realistic faces**
Music Generation	**Drug Discovery**	**Game Content**
AI composers, style transfer	*Generate new molecule designs*	*Generate textures, levels, characters*
Create new songs in any style!	**Design new medicines faster**	**Create infinite game content!**

Three Main Types We'll Cover
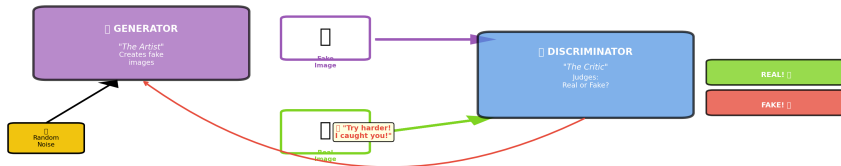
1. **GANs (Generative Adversarial Networks):** Two networks compete to create realistic images

GANs: Generative Adversarial Networks

GANs: The Artist vs The Critic Game ✕

🤔 HOW THE "GAME" WORKS:

🎨 THE GENERATOR (Artist):
• Starts with random noise (like random brushstrokes)
• Creates fake images trying to look real
• Goal: "Fool the critic into thinking my art is real!"

🔍 THE DISCRIMINATOR (Critic):
• Looks at both real images and fake images
• Tries to tell them apart
• Goal: "Spot the fakes! Don't be fooled!"

🏆 THE COMPETITION:
1️⃣ Generator creates a fake image
2️⃣ Discriminator judges: "Real or Fake?"
3️⃣ If caught: Generator learns from mistakes and improves
4️⃣ Discriminator also gets better at detecting fakes
5️⃣ They keep competing until Generator is SO GOOD, even the expert Discriminator can't tell!

🎉 THE RESULT:
After thousands of rounds, the Generator becomes an AMAZING artist!
It can create super realistic images that never existed before!

🌍 REAL-WORLD ANALOGY:
Think of it like a student artist (Generator) trying to forge a painting,
while an art expert (Discriminator) tries to detect forgeries.
The student keeps improving until they're as good as a master!

💡 Why GANs are MAGICAL:
• They learn to create realistic images WITHOUT being told exactly how
• The competition makes both networks better and better
• No human has to teach them what "realistic" means - they figure it out!

Simple Explanation

Two neural networks compete:

- **Generator:** Creates fake images (like an art forger)

Artbreeder

What it does:

- Generate unique portraits
- Mix different faces together
- Adjust age, gender, ethnicity
- Create landscapes, album covers
- Used by 10+ million users

How Artists Use It

- Book cover illustrations
- Character design for games
- Concept art for films
- Social media content

ThisPersonDoesNotExist.com

- Generates random faces
- 100% synthetic people
- Photorealistic quality
- New face every refresh
- Built with StyleGAN

Try It Yourself!

Visit the website - every face you see was created by AI, not a photo!

GAN Application: Deepfake Detection

The Problem

Malicious Uses:

- Fake celebrity videos
- Misinformation campaigns
- Identity fraud
- Non-consensual content

The Solution

GANs fight GANs:

- Train detectors on fake data
- Identify artifacts and inconsistencies
- Real-time video verification
- Protect public figures

Real Deployments

- **Facebook/Meta:** Deepfake detection system
- **Microsoft:** Video Authenticator tool
- **Intel:** FakeCatcher (96% accuracy)
- **Adobe:** Content Authenticity Initiative

Arms Race

Detection technology must constantly evolve as GANs improve!

Why Generate Medical Data?

Privacy & Scarcity Issues:

- Real patient data is private (HIPAA)
- Rare diseases lack training samples
- Hard to share data between hospitals
- Need diverse examples for AI training

What GANs Generate

- Synthetic X-rays
- Artificial MRI scans
- Fake patient records
- Privacy-preserving datasets

Real Research Applications

- **Mayo Clinic:** Generate rare tumor samples
- **Stanford:** Synthetic chest X-rays
- **MIT:** Privacy-safe medical records
- Train better AI without compromising privacy

Impact

Enables medical AI research while protecting patient privacy!

Modern Game Development

How GANs Help:

- Generate unique NPC faces
- Create diverse character variations
- Design textures and materials
- Procedural content generation
- Speed up asset creation

Real Game Studios

- **EA Sports:** Generate realistic player faces
- **Ubisoft:** NPC diversity in Assassin's Creed
- Reduce manual art time by 70%

Player Customization

- Infinite character appearance options
- Realistic face generation
- Upload photo for custom avatar
- AI-assisted character design

Industry Adoption

Major game engines (Unity, Unreal) now integrate GAN-based tools!

AI Fashion Designers

What They Generate:

- New clothing designs
- Pattern and texture variations
- Color scheme combinations
- Style transfer between eras
- Personalized recommendations

Fashion Companies Using AI

- **Stitch Fix:** Personalized designs
- **Tommy Hilfiger:** IBM collaboration
- **Zalando:** Generated fashion models

Virtual Try-On

- Generate how clothes look on you
- Try outfits without physically wearing
- Reduce online shopping returns
- Personalized styling suggestions

Business Impact

AI-designed collections sell out 30% faster than traditional designs!
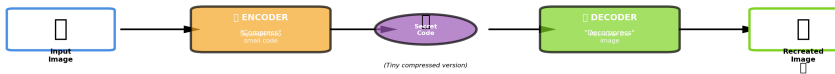
VAEs: Variational Autoencoders

VAE: Learning to Compress and Create

VAE: The Compression Artist

```
┌──────────┐      ┌── ENCODER ──┐       ╭── Secret ──╮      ┌── DECODER ──┐      ┌──────────┐
│  Input   │ ───► │ "Compress"  │ ───►  │   Code     │ ───► │"Decompress" │ ───► │ Recreated│
│  Image   │      │ small code  │       ╰────────────╯      │   image     │      │  Image   │
└──────────┘      └─────────────┘     (Tiny compressed     └─────────────┘      └──────────┘
   Input                                   version)                                Recreated
   Image                                                                            Image
```

VAE IN SIMPLE TERMS (Variational Autoencoder):

THINK OF IT LIKE ZIP FILES:
When you zip a large file, it gets compressed into something smaller.
VAEs do the same thing with images, but in a SMART way!

HOW IT WORKS:

1️⃣ ENCODER (The Compressor):
• Takes a big image (like a photo of a cat)
• Squeezes it into a tiny "secret code" (just a few numbers!)
• Like compressing a photo to save space on your phone

2️⃣ LATENT SPACE (The Secret Code):
• The compressed representation
• Contains the ESSENCE of the image (not every pixel, just the important info)
• Like a recipe card instead of the full meal

3️⃣ DECODER (The Decompressor):
• Takes the secret code
• Recreates the image from memory
• Like unzipping a file, but it GENERATES the image

THE MAGIC TRICK:
Once trained, you can:
Create NEW images by giving it NEW random codes!
Mix codes to blend images (cat + dog = ???)
Change codes slightly to modify images

REAL-WORLD ANALOGY:
Imagine describing a face with just a few words: "round face, blue eyes, blonde hair, smiling"
From just that SHORT description, someone could draw a complete face!
That's what VAE's secret code does!

VAE vs GAN:
• VAE: More stable, creates smooth variations, great for compression
• GAN: More realistic, but training is trickier (needs the competition)

USES:
• Generating new faces, artwork, designs
• Image compression (better than JPEG!)
• Anomaly detection (if compression fails, something's weird!)
• Creating variations of existing images

Simple Explanation

VAEs compress data into a small code, then decompress it:

Manufacturing Quality Control

How It Works:

- Train VAE on normal products
- VAE learns what "normal" looks like
- Defects reconstruct poorly
- High reconstruction error = defect!

Real Applications

- Detect scratches on surfaces
- Find cracks in materials
- Identify missing components
- Automated quality inspection

Other Anomaly Detection Uses

- **Cybersecurity:** Detect network intrusions
- **Finance:** Identify fraudulent transactions
- **Healthcare:** Flag unusual patient vitals
- **IoT:** Detect sensor failures

Advantage

Works without labeled defect examples - learns from normal data only!

VAE Application: Image Compression

Why VAEs for Compression?

Advantages over JPEG:

- Better quality at low bitrates
- Learned compression (adapts to content)
- Can compress to tiny sizes
- Semantic preservation

How It Works

- Encoder compresses to latent code
- Store only the small code
- Decoder reconstructs when needed
- 10-100x smaller than JPEG

Real-World Uses

- Store medical imaging archives
- Stream video at lower bandwidth
- Compress satellite imagery
- Mobile app image caching

Research Example

Google's neural image compression beats JPEG by 50% in quality metrics!

VAE Application: Drug Molecule Generation

Pharmaceutical Discovery

Traditional Approach:

- Test millions of molecules
- Takes 10+ years per drug
- Costs billions of dollars
- High failure rate

VAE Approach

- Learn from existing drugs
- Generate similar molecules
- Optimize for target properties
- Find candidates much faster

Real Pharmaceutical AI

- **Insilico Medicine:** Generated novel molecules
- **Atomwise:** AI drug discovery platform
- **BenevolentAI:** COVID-19 drug repurposing
- Reduce discovery time by 75%
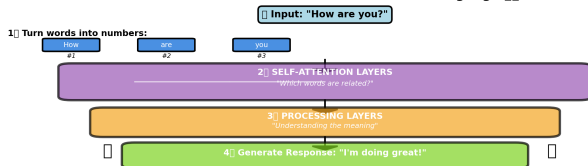
Major Milestone

First AI-discovered drug entered human trials in 2020!

Transformers: The Revolution in AI

Transformers: How ChatGPT Understands Language

Input: "How are you?"

1️⃣ Turn words into numbers:

How	are	you
#1	#2	#3

2️⃣ SELF-ATTENTION LAYERS
"Which words are related?"

3️⃣ PROCESSING LAYERS
"Understanding the meaning"

4️⃣ Generate Response: "I'm doing great!"

TRANSFORMER IN SIMPLE TERMS:

OLD APPROACH (RNNs): Read sentence word-by-word, left to right
• Like reading a book one word at a time
• Slow and forgets earlier words
• Example: By the time you read "dog", you forgot the "big brown" at the start

TRANSFORMER APPROACH: Look at ALL words at once!
• Like looking at the whole sentence together
• Uses ATTENTION to see which words are related
• Fast and remembers everything

THE MAGIC - SELF-ATTENTION:
When processing the word "it", the transformer looks at the whole sentence:
"The cat sat on the mat because it was tired"
→ Attention figures out "it" refers to "cat", not "mat"!

HOW CHATGPT USES THIS:
1️⃣ You type: "Translate 'hello' to Spanish"
2️⃣ Transformer reads ALL words simultaneously
3️⃣ Attention connects "translate" with "hello" and "Spanish"
4️⃣ It understands: [ACTION: translate] [FROM: hello] [TO: Spanish]
5️⃣ Generates: "Hola"

WHY TRANSFORMERS CHANGED AI:
✓ Can be trained on MASSIVE amounts of text (entire internet!)
✓ Understands context better (knows "bank" = money vs river)
✓ Much faster than old methods (processes sentences in parallel)
✓ Can handle very long text (remember things from paragraphs ago)

KEY INSIGHT:
Transformers are like having a super-smart friend who can:
• Read an entire book in seconds
• Remember every detail
• Understand how everything connects
• Explain it back to you in simple terms

Simple Explanation

Transformers process sequences by paying attention to relevant parts:

• Designed for text, but work on images/audio too

Transformers in Daily Life: You Use These Every Day!

🌐 Language Translation

Google Translate, DeepL

Translate between 100+ languages!

💬 Chatbots

ChatGPT, Claude, Gemini

Have natural conversations!

✍️ Writing Assistants

Grammarly, Notion AI, Writing tools

Help write better emails & docs!

💻 Code Generation

GitHub Copilot, ChatGPT coding

Write code from descriptions!

📝 Text Summarization

Summarize articles, papers, books

TLDR: Get the main points fast!

❓ Question Answering

Search engines, Virtual assistants

Answer questions from documents!

Why Transformers Changed Everything

Before 2017: RNNs struggled with long sequences. After 2017: Transformers enabled GPT, BERT, and the current AI revolution!

What ChatGPT Can Do

Capabilities:

- Answer questions
- Write code and debug
- Compose essays and emails
- Explain complex topics
- Translate languages
- Creative writing

Real Usage Statistics

- 100+ million weekly users
- Fastest-growing consumer app
- Used in 185+ countries

How Students Use It

- Homework help and tutoring
- Research assistance
- Programming debugging
- Study guide creation
- Language learning
- Career advice

Built With Transformers

GPT-4 uses a massive transformer with 175+ billion parameters!

Old vs New Approach

Before Transformers (2016):

- Phrase-based translation
- Limited context understanding
- Often awkward output

After Transformers (2017+):

- Sentence-level context
- Natural, fluent translations
- 60% reduction in errors

Features Powered by Transformers

- 133 languages supported
- Real-time conversation mode
- Camera translation (point and translate)
- Offline translation
- Context-aware results

Daily Impact

500+ million people use Google Translate every day!

AI Pair Programmer

What Copilot Does:

- Suggests code as you type
- Writes entire functions
- Explains existing code
- Converts comments to code
- Generates tests
- Fixes bugs

Real Developer Impact

- 46% of code written by AI
- 55% faster task completion
- Used by 1.2 million developers

How It Works

- Built on GPT (Codex model)
- Trained on billions of lines of code
- Understands context from your files
- Suggests in real-time
- Supports 12+ programming languages

For Students

Great learning tool - see how experts solve problems!

Gmail Smart Compose

Features:

- Suggests next words/sentences
- Learns your writing style
- Adapts to context
- Multi-language support
- Works on mobile too

Time Savings

- Average user saves 1 billion characters/week
- Reduces writing time by 11%
- 4+ billion emails use it daily

Other Email AI Features

- **Smart Reply:** Suggest full responses
- **Subject suggestions:** Auto-generate subjects
- **Tone adjustment:** Make emails more formal
- **Grammar correction:** Fix mistakes

All Powered by Transformers

These "small" conveniences use the same tech as ChatGPT!

Automatic Summarization

What It Does:

- Read long documents
- Extract key points
- Generate concise summary
- Preserve important details
- Save reading time

Real Products

- **Microsoft Word:** Auto-summarize
- **Slack:** Thread summaries
- **Notion AI:** Note summarization
- **Chrome extensions:** Web page summaries

Use Cases

- Research paper summaries
- News article digests
- Legal document review
- Meeting notes condensation
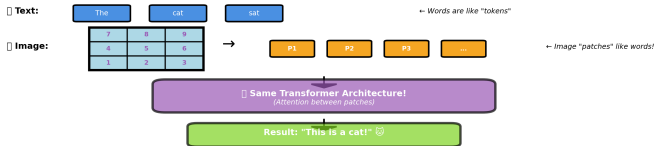- Customer feedback analysis

Productivity Boost

Lawyers using AI summarization save 60% of document review time!

Vision Transformers: Teaching Language Models to See!

THE BIG IDEA: Treat images like sentences!

Text: The | cat | sat   ← Words are like "tokens"

Image: → P1 | P2 | P3 | ...   ← Image "patches" like words!

Same Transformer Architecture!
(Attention between patches)

Result: "This is a cat!" ☺

VISION TRANSFORMERS (ViT) IN SIMPLE TERMS:

THE CLEVER TRICK:
Instead of inventing a NEW architecture for images, we realized:
"Why not use the SAME transformer that works great for text?"

HOW IT WORKS:
1. Cut image into small squares (patches) - like cutting a pizza into slices
2. Flatten each patch into a list of numbers (treat it like a "word")
3. Feed these "image words" into a regular transformer
4. Let attention figure out which patches are related!

EXAMPLE:
If you have a photo of a cat:
• Patch 1: Contains part of the ear
• Patch 2: Contains the eye
• Patch 3: Contains whiskers
→ Attention learns: "These patches together = cat face!"

ViT vs CNN:
• CNN: Looks at local neighborhoods (like looking through a magnifying glass)
• ViT: Looks at the whole image at once (like seeing the big picture)
• ViT: Can find connections between far-apart parts of the image!

WHY IT'S AWESOME:
✓ Simpler architecture than CNNs (one model for everything!)
✓ Can handle huge images (just use more patches)
✓ Learns better from LOTS of data
✓ Same code works for images, text, video, audio!

THE FUTURE:
Multi-modal models like GPT-4, Claude use this idea:
→ They can understand BOTH text AND images together!
→ Ask: "What's in this photo" and show a picture
→ AI can answer because it treats images and text the same way!

Vision Transformers (ViT)

Applying transformers to images:

- Break image into patches (like words)
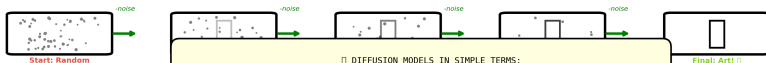
Diffusion Models: The Newest Revolution

Diffusion Models: The Magic of "Denoising"

→ FORWARD PROCESS: Gradually Add Noise (Training)

Clean Image → Step 1 → Step 2 → Step 3 → Pure Noise

→ REVERSE PROCESS: Gradually Remove Noise (Generation)

Start: Random → → → → Final: Art!

DIFFUSION MODELS IN SIMPLE TERMS:

THE CORE IDEA: Learn to reverse the process of adding noise!

ANALOGY - Like a Magic Eraser:
Imagine you have a beautiful drawing, but someone scribbled all over it with random marks.
A diffusion model is like a magic eraser that can gradually remove the scribbles
until the original drawing appears again!

TWO PHASES:

1️⃣ TRAINING (Forward): Add noise step by step
• Start with real images (like photos of cats)
• Gradually add random noise until it's completely fuzzy
• Save each step to learn the pattern

2️⃣ GENERATION (Reverse): Remove noise step by step
• Start with pure random noise (like TV static)
• AI predicts: "What does this look like with LESS noise?"
• Remove a little noise at a time
• Eventually reveals a clear image!

HOW THE AI LEARNS:
During training, the AI learns: "Given a noisy image, what's the clean version?"
It practices this over and over with millions of images.
Then during generation, it applies this skill repeatedly to create new images!

WHY IT'S SPECIAL:
• Very stable training (unlike GANs that can be tricky)
• High quality results (super realistic images!)
• Controllable (you can guide what it creates with text)
• Flexible (works for images, audio, video, 3D models)

THE MAGIC:
You start with random noise (complete chaos) and step-by-step, the AI
sculpts it into a beautiful image - like a sculptor revealing a statue
hidden inside a block of marble!

Generative Models Comparison: Which One to Use?

	GANs		VAEs		Diffusion
	(Generator vs Discriminator)		(Encoder-Decoder)		(Iterative Denoising)
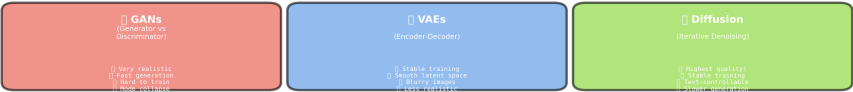
DETAILED COMPARISON:

Feature	GANs	VAEs	Diffusion
Image Quality			
Training Stability			
Generation Speed			
Diversity			
Control (text)			

WHEN TO USE EACH:

GANs: When you need fast generation and have lots of training data
Example: Generating realistic faces, style transfer

VAEs: When you need smooth interpolation and stable training
Example: Data compression, anomaly detection, simple generation

DIFFUSION: When you need the highest quality and text control
Example: AI art (DALL-E, Midjourney), image editing, creative applications

CURRENT WINNER: Diffusion Models!
They're the technology behind most modern AI art tools because they:
• Generate the highest quality images
• Work great with text prompts
• Are stable and reliable to train
• Can be controlled and edited easily

THE TREND:
Diffusion models have largely replaced GANs for image generation!
They're slower, but the quality difference is worth it for most applications.

GANs

Pros: Fast generation

VAEs

Pros: Stable, good latent space

Diffusion

Pros: Best quality, stable

Diffusion Models: Creating the Impossible!

AI Art Creation
Midjourney, DALL-E 3, Stable Diffusion

Create stunning art from text!

Image Editing
Inpainting, outpainting, style transfer

Edit photos like a pro designer!

Video Generation
Runway, Pika Labs, animation tools

Create videos from text!

Architecture & Design
Generate room layouts, building designs

Design spaces instantly!

Game Assets
Generate textures, characters, environments

Create game art super fast!

Fashion & Product
Design clothes, product mockups

Prototype designs in seconds!

Why Diffusion Models Won

They power DALL-E 2, Midjourney, Stable Diffusion - the best AI image generators today!

What DALL-E 2 Can Do

Text-to-Image Generation:

- Type a description, get an image
- Photorealistic or artistic styles
- Combine multiple concepts
- Edit existing images
- Outpainting (extend images)

Example Prompts

- "A cat astronaut on Mars"
- "Oil painting of a sunset over Manila"
- "Teddy bear shopping for groceries"

Real-World Uses

- Marketing content creation
- Concept art for entertainment
- Educational illustrations
- Social media graphics
- Product mockups

By OpenAI

Same company behind ChatGPT - 1.5+ million users create images daily!

What Makes Midjourney Special

Artistic Focus:

- Exceptionally beautiful outputs
- Strong artistic style
- Great for fantasy/sci-fi art
- Discord-based interface
- Community of 16+ million users

Popular Use Cases

- Book cover designs
- Album artwork
- Game concept art
- NFT art generation

Industry Impact

- Artists use it for inspiration
- Magazine covers created with AI
- Award-winning art competitions
- Commercial illustration work

Controversy

AI art won Colorado State Fair - sparked debate about AI creativity!

Diffusion Application: Stable Diffusion

Why Stable Diffusion is Different

Open Source:

- Free to use and modify
- Run on your own computer
- Customize and fine-tune
- No usage restrictions
- Active developer community

Technical Details

- Can run on consumer GPUs
- Faster than DALL-E 2
- Extensible with plugins
- Multiple versions and variants

Popular Applications Built With It

- DreamStudio (official interface)
- Automatic1111 (popular UI)
- ComfyUI (node-based editor)
- Mobile apps (Draw Things)
- Photoshop plugins

Democratizing AI

Anyone with a decent computer can now generate professional-quality images!

Professional Image Editing

Firefly Features:

- Text-to-image generation
- Generative fill (edit parts of images)
- Text effects (3D text styles)
- Generative recolor
- Integrated in Photoshop

Key Advantages

- Trained on Adobe Stock (licensed data)
- Commercially safe to use
- Professional quality outputs
- Seamless Creative Cloud integration

Real Designer Workflows

- Remove unwanted objects
- Extend backgrounds
- Generate variations quickly
- Create mockups from descriptions
- Speed up creative process 10×

Industry Standard

Adobe's AI tools are becoming essential for professional designers!

Text-to-Video AI

Emerging Applications:

- Generate short video clips
- Animate static images
- Create transitions
- Style transfer for video
- AI-assisted editing

Current Platforms

- **Runway Gen-2:** Text-to-video
- **Pika Labs:** Video generation
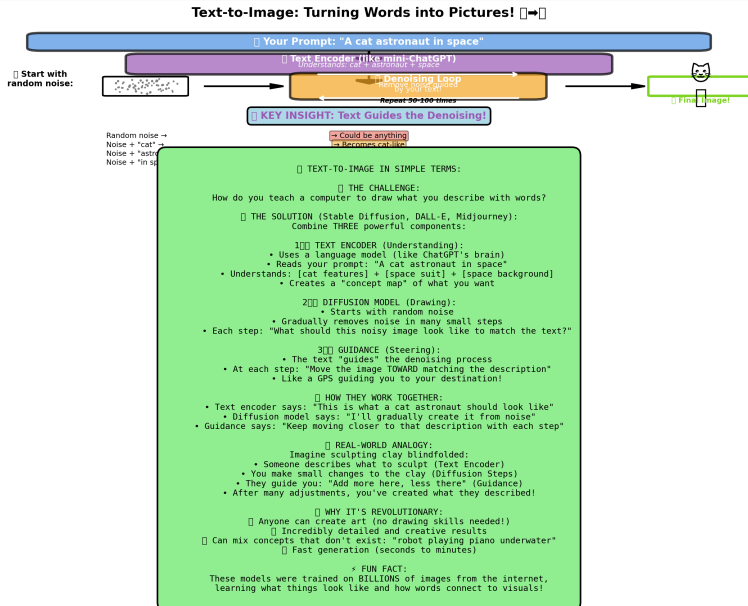- **Stable Video Diffusion:** Open source

Use Cases

- Social media content
- Marketing videos
- Animated presentations
- Film pre-visualization
- Game cinematics

Future is Coming

Video generation is improving rapidly - expect major breakthroughs soon!

Text-to-Image: Turning Words into Pictures! 📝➡️🖼️

🗣️ Your Prompt: "A cat astronaut in space"

🧠 Text Encoder (like mini-ChatGPT)
Understands: cat + astronaut + space

🎲 Start with
random noise:

Denoising Loop
by Wild Tech
Repeat 50-100 times

🖼️ Final Image!

🔑 KEY INSIGHT: Text Guides the Denoising!

Random noise →
Noise + "cat" →
Noise + "astr
Noise + "in sp

→ Could be anything
→ Becomes cat-like

🎨 TEXT-TO-IMAGE IN SIMPLE TERMS:

🤔 THE CHALLENGE:
How do you teach a computer to draw what you describe with words?

💡 THE SOLUTION (Stable Diffusion, DALL-E, Midjourney):
Combine THREE powerful components:

1️⃣ TEXT ENCODER (Understanding):
• Uses a language model (like ChatGPT's brain)
• Reads your prompt: "A cat astronaut in space"
• Understands: [cat features] + [space suit] + [space background]
• Creates a "concept map" of what you want

2️⃣ DIFFUSION MODEL (Drawing):
• Starts with random noise
• Gradually removes noise in many small steps
• Each step: "What should this noisy image look like to match the text?"

3️⃣ GUIDANCE (Steering):
• The text "guides" the denoising process
• At each step: "Move the image TOWARD matching the description"
• Like a GPS guiding you to your destination!

🔄 HOW THEY WORK TOGETHER:
• Text encoder says: "This is what a cat astronaut should look like"
• Diffusion model says: "I'll gradually create it from noise"
• Guidance says: "Keep moving closer to that description with each step"

🎨 REAL-WORLD ANALOGY:
Imagine sculpting clay blindfolded:
• Someone describes what to sculpt (Text Encoder)
• You make small changes to the clay (Diffusion Steps)
• They guide you: "Add more here, less there" (Guidance)
• After many adjustments, you've created what they described!

🌟 WHY IT'S REVOLUTIONARY:
🎨 Anyone can create art (no drawing skills needed!)
🖼️ Incredibly detailed and creative results
🔀 Can mix concepts that don't exist: "robot playing piano underwater"
⚡ Fast generation (seconds to minutes)

✨ FUN FACT:
These models were trained on BILLIONS of images from the internet,
learning what things look like and how words connect to visuals!

How It All Works Together

1. **Text Encoder (Transformer):** Understand your description

Ethics & Responsible AI

⚖ **Ethics: The Responsible Use of AI Generation**

✓ POSITIVE USES

⚠ CONCERNS & RISKS

🎨 **Creative Expression**	*Artists using AI as a tool*
🔬 **Medical Research**	*Generating synthetic medical data*
📚 **Education**	*Creating learning materials*
♿ **Accessibility**	*Helping people with disabilities*
🔭 **Scientific Discovery**	*Simulating molecules, proteins*

🎭 **Deepfakes**	*Fake videos spreading misinformation*
©️ **Copyright Issues**	*Who owns AI-generated art?*
💼 **Job Displacement**	*Will AI replace human artists?*
🔓 **Privacy Violations**	*Generating fake images of real people*
📰 **Fake News**	*Creating fake but realistic content*

🧭 HOW TO USE GENERATIVE AI RESPONSIBLY:

1️⃣ BE TRANSPARENT: Always disclose when content is AI-generated
Example: Label AI art as "Created with AI" or "AI-assisted"

2️⃣ RESPECT CONSENT: Don't create fake images/videos of real people without permission
Example: Don't deepfake someone's face onto inappropriate content

3️⃣ VERIFY SOURCES: Don't believe everything you see online
Example: Check if that viral video might be AI-generated

4️⃣ CREDIT PROPERLY: Acknowledge both the AI tool AND human creativity
Example: "Concept by [Artist], generated with [AI Tool]"

5️⃣ USE FOR GOOD: Think about the impact of what you create
Example: Use AI to help people, not to deceive or harm them

💡 REMEMBER: Just because we CAN create something doesn't mean we SHOULD!

🎯 THE GOAL: Use AI as a tool to enhance human creativity and solve problems,
not to deceive, harm, or replace human connection.

Important Questions to Consider

As these technologies become powerful, we must think carefully about their impact!

Key Ethical Issues

Misinformation & Deepfakes

Concerns:

- Fake news and propaganda
- Identity fraud
- Non-consensual content
- Erosion of trust in media

Solutions:

- Detection technology
- Digital watermarking
- Media literacy education
- Legal frameworks

Bias & Fairness

Problems:

- Biased training data
- Perpetuating stereotypes
- Unfair representation
- Discrimination in outputs

Mitigation:

- Diverse training datasets
- Bias testing and auditing
- Responsible AI guidelines
- Inclusive development teams

Copyright & Intellectual Property

Questions:

- Who owns AI-generated content?
- Is training on copyrighted data fair use?
- Should artists be compensated?
- How to attribute AI creations?

Current Debates:

- Ongoing lawsuits (artists vs AI companies)
- New legislation being proposed
- Industry opt-out mechanisms

Job Displacement

Concerns:

- Will AI replace creative jobs?
- Impact on artists, writers, designers
- Economic inequality
- Need for reskilling

Opportunities:

- AI as a tool, not replacement
- New creative possibilities
- Democratization of creation
- Focus on uniquely human skills

Your Responsibility

As future AI practitioners, think critically about the impact of your work!

Key Concept: Attention Mechanism

Attention Mechanism: Focusing on What Matters

OLD WAY: Reading Every Word Equally

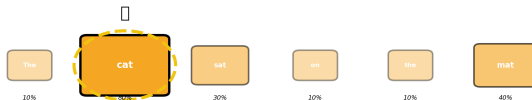| The | cat | sat | on | the | mat |

Problem: Computer treats ALL words as equally important! ☹

Question: "What sat on the mat?" → Hard to know "cat" is the answer!

NEW WAY: Attention! Focus on Important Words

| The | cat | sat | on | the | mat |
| 10% | 30% | 30% | 10% | 10% | 40% |

Attention lets the model FOCUS on important words!

ATTENTION IN SIMPLE TERMS:

When you read "The cat sat on the mat," you naturally focus more on "cat" and "mat" because they're the important nouns. Words like "the" and "on" are less important.

ATTENTION teaches AI to do the same thing - focus on what matters!

What is Attention?

A mechanism that lets neural networks focus on relevant parts:

- In text: Focus on important words in a sentence

Problem Without Attention

Translating: "The cat sat on the mat"

Old approach:

- Process word by word left to right
- Forget earlier context
- Struggle with long sentences
- Poor word alignment

With Attention

For each output word, the model:

- Looks at ALL input words
- Focuses on relevant ones
- "sat" pays attention to "cat" and "mat"
- Handles long-distance dependencies
- Better translation quality

Why It's Revolutionary

Attention enabled Transformers to outperform all previous architectures!

Using These Models: Practical Guide

Free/Accessible Tools

Try these today:

- **ChatGPT:** Free tier available
- **Bing Image Creator:** Free DALL-E access
- **Google Colab:** Run Stable Diffusion free
- **Hugging Face:** Try many models online
- **Runway:** Free trial for video

For Developers

Build your own:

- PyTorch or TensorFlow
- Hugging Face Transformers library
- Stable Diffusion on GitHub
- Pre-trained models available
- Fine-tune on your data

Learning Resources

- Fast.ai courses (free)
- Hugging Face tutorials
- Papers with Code
- YouTube: Two Minute Papers

Start Small

Use existing models before building from scratch - learn by doing!

Writing Good Prompts

Be specific:

- Describe style (photorealistic, cartoon, oil painting)
- Specify details (colors, lighting, mood)
- Mention composition (close-up, wide shot)
- Add quality keywords (4K, detailed, masterpiece)

Example Good Prompt

"A majestic golden retriever sitting in a flower meadow at sunset, photorealistic, warm lighting, shallow depth of field, 4K quality"

Iteration is Key

- Generate multiple variations
- Refine your prompt
- Use negative prompts (what to avoid)
- Adjust parameters (steps, guidance)
- Learn from community prompts

Pro Tip

Check out prompt libraries (Lexica.art, PromptHero) to learn from others!

Challenge: Poor Results

If outputs look bad:

- Improve your prompt specificity
- Try different seed values
- Adjust generation parameters
- Use a different model/variant
- Increase generation steps

Challenge: Wrong Anatomy/Details

Known limitations:

- Hands and fingers often wrong
- Text in images unclear
- Physics may be incorrect
- Use inpainting to fix specific parts

Challenge: Slow Generation

Speed up:

- Use lower resolution first
- Reduce number of steps
- Try faster samplers
- Use GPU acceleration
- Consider paid services for speed

Challenge: Reproducibility

Get consistent results:

- Save your seed numbers
- Keep prompt exactly the same
- Note all parameters used
- Use img2img for variations

Summary & Looking Forward

What We Learned

Five major architectures changing the world:

1. **CNNs:** Revolutionized computer vision (medical imaging, self-driving cars, face recognition)
2. **GANs:** Generate realistic images (AI art, deepfakes, synthetic data)
3. **VAEs:** Compress and generate (anomaly detection, drug discovery)
4. **Transformers:** Dominated NLP (ChatGPT, translation, code generation)
5. **Diffusion:** Best image generation (DALL-E 2, Midjourney, Stable Diffusion)

Main Message

These aren't just research projects - they're tools you can use TODAY in real applications!

CNNs Applications

- Medical tumor detection
- Self-driving lane detection
- Phone face unlock
- Security cameras
- Satellite imagery analysis

GAN Applications

- Artbreeder AI art
- Deepfake detection
- Synthetic medical data
- Game character creation
- Fashion design

Transformer Applications

- ChatGPT conversations
- Google Translate
- GitHub Copilot
- Email auto-complete
- Document summarization

Diffusion Applications

- DALL-E 2 image generation
- Midjourney art creation
- Stable Diffusion (open source)
- Adobe Firefly editing
- Video generation (emerging)

Trends to Watch

Next 1-2 years:

- **Multimodal AI:** Text, image, audio, video together
- **Better video generation:** Movie-quality AI videos
- **3D generation:** Create 3D models from text
- **Real-time generation:** Instant results
- **Personalization:** AI that learns your style

Career Opportunities

Skills in demand:

- AI/ML engineering
- Prompt engineering
- AI safety and ethics
- Creative AI applications
- AI product management

Get Involved

The best way to learn is to experiment - start building today!

Hands-On Practice

- Try Stable Diffusion on Colab
- Build projects with Hugging Face
- Fine-tune models on your data
- Participate in Kaggle competitions
- Contribute to open source projects

Online Courses

- Fast.ai: Practical Deep Learning
- Stanford CS230: Deep Learning
- Coursera: Deep Learning Specialization
- Hugging Face NLP Course (free)

Stay Updated

- Follow Papers with Code
- Read AI newsletters (The Batch, etc.)
- Watch Two Minute Papers (YouTube)
- Join AI Discord communities
- Attend local meetups

Next Steps in This Course

Workshop: Hands-on coding with ResNet, GPT-2, Stable Diffusion - let's use these models!

Thank you for your attention!

Contact Information

Instructor: Noel Jeffrey Pinton
Course: CMSC 173 - Machine Learning
Institution: University of the Philippines - Cebu
Department: Computer Science

Remember

Advanced neural networks are tools that empower creativity and solve real problems. Use them responsibly and ethically!