Генеративті нейрондық желілер және Transfer Learning

Деректерді генерациялау және үйренілген білімді қайта пайдалану

Дәрістің негізгі міндеттері

- Дискриминативті және генеративті модельдерді ажырату.
- VAE және GAN құрылымдары мен оқыту мақсаттарын түсіндіру.
- Feature extraction және fine-tuning стратегияларын негіздеу.

Дискриминативті және генеративті модельдер

Дискриминативті модель (Discriminative Model)

Кіріс объекті x берілген кезде оның класы немесе мақсатты мәні y анықталады.

Кіріс дерек x
(жиі өлшемді, мысалы, 28 × 28)



$x \in \mathbb{R}^n$
$(n = 784)$

Дискриминативті модель
$f_\theta(x)$

Шығыс y
(класс немесе мақсатты мән)

$P(y = 0 \mid x)$	0.01
$P(y = 1 \mid x)$	0.02
$P(y = 2 \mid x)$	0.05
$P(y = 3 \mid x)$	0.90
$P(y = 4 \mid x)$	0.02
...	

$$\hat{y} = \arg\max_y P(y \mid x) = 3$$

Үйренетін тәуелділік:

$$p(y \mid x) \qquad (9.1)$$

Берілген x объектісі үшін бенгілі бір y класын (немесе мәнін) болжау.

Генеративті модель (Generative Model)

Деректердің таралуын $p(x)$ үйреніп, осы таралудан жаңа объектілер қалыптастырады.

Жасырын айнымалы z
(төмен өлшемді)

$z \sim p(z)$

$$\begin{bmatrix} 0.3 \\ -1.2 \\ 0.8 \\ \vdots \end{bmatrix}$$

$z \in \mathbb{R}^d$
$(d \ll n)$

Генеративті модель
$G_\theta(z)$

Генерацияланған дерек x_{new}
(жаңа объект)



$x_{new} \in \mathbb{R}^n$
$(n = 784)$

Үйренетін таралу:

$$p(x) \qquad (9.2)$$

Деректердің жалпы таралуын модельдеу және осы таралудан жаңа x объектілерін генерациялау.

Сипаттама	Дискриминативті модель	Генеративті модель
Негізгі мақсат	Берілген x үшін y-ды болжау	Деректер таралуын үйрену және жаңа x объектілерін генерациялау
Үйренетін тәуелділік	Шартты ықтималдық $p(y \mid x)$	Деректер таралуы $p(x)$ немесе $p(x \mid y)$
Кіріс және шығыс	$x \to y$	$z \to x_{new}$
Жиі қолданылатын есептер	Классификация, регрессия	Кескін, мәтін, дыбыс генерациясы, деректерді толықтыру, аномалияларды модельдеу
Мысал модельдер	MLP, CNN, RNN, т.б.	VAE, GAN, диффузиялық модельдер, автогрегрессивті модельдер, т.б.

Дереккөз: 9-дәріс конспектісі, 2-бет, 9.1-сурет.

Генеративті оқыту: таралуды модельдеу

$$p_{model}(x) \approx p_{data}(x)$$

Дискриминативті мақсат

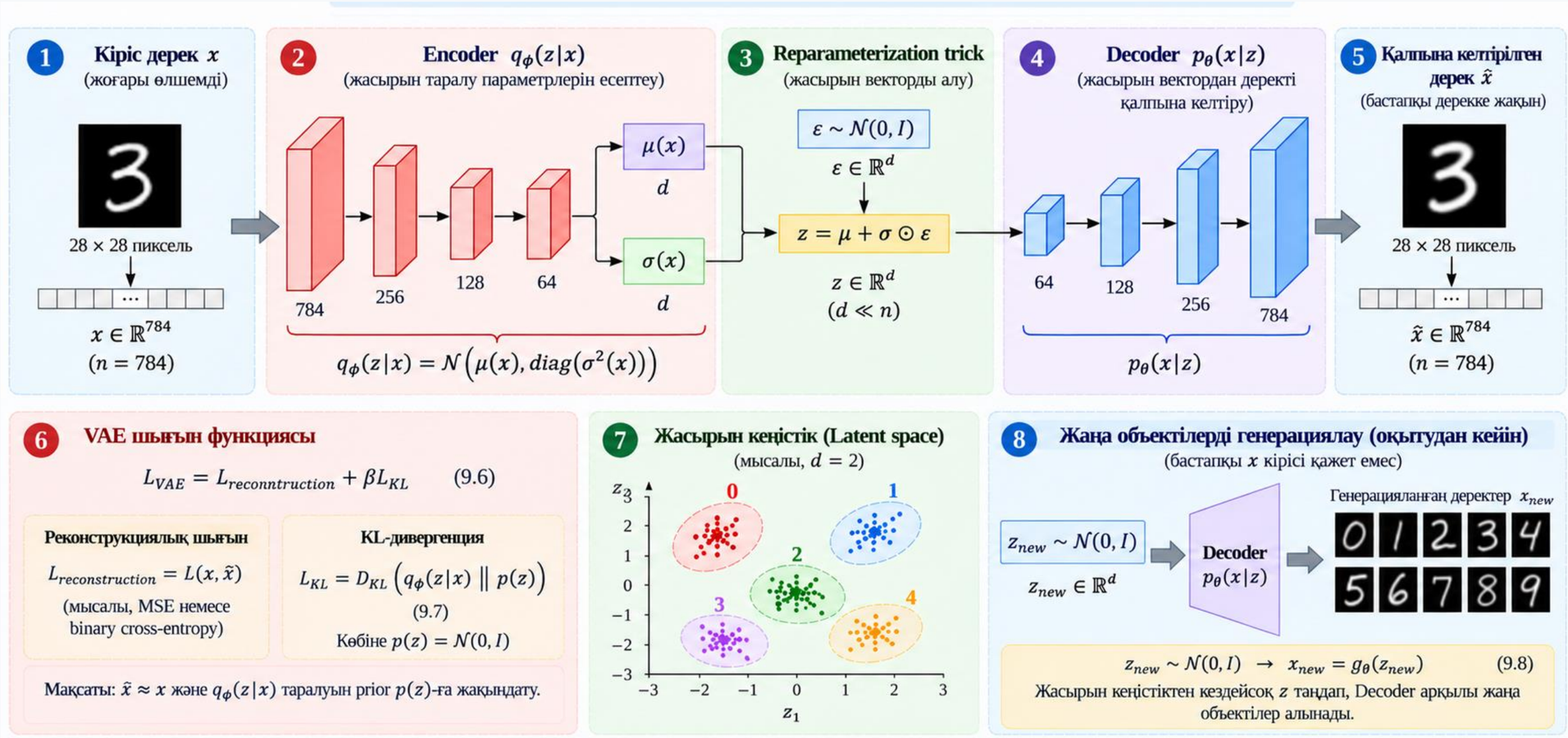
p(y | x): берілген объект үшін класс не мақсатты мәнді болжау.

Генеративті мақсат

p(x) немесе p(x | y): деректер таралуын не шартты таралуды модельдеу.

Жасырын айнымалысы бар үлгі: z ~ p(z) → G(z) → x_жаңа.

Variational Autoencoder — VAE



Суреттегі latent кластерлер — түсіндірмелік үлгі; олардың айқын бөлінуіне кепілдік жоқ.

Дереккөз: конспект, 3-бет, 9.2-сурет.

VAE: қайта параметрлеу және шығын

$$q_\varphi(z \mid x) = \mathcal{N}(\mu(x), \mathrm{diag}(\sigma^2(x)))$$

$$z = \mu + \sigma \odot \varepsilon,\ \varepsilon \sim \mathcal{N}(0, I)$$

$$L_{VAE} = L_{reconstruction} + \beta\, D_{KL}(q_\varphi(z|x) \,||\, p(z))$$

Reconstruction — кірісті қалпына келтіру сапасы.
KL — жуық posterior-ды prior $p(z) = \mathcal{N}(0, I)$-ға жақындату.
$\beta = 1$ — стандартты VAE; β өзгерсе, осы тепе-теңдік өзгереді.

VAE арқылы жаңа үлгіні генерациялау

Оқыту кезінде

x → Encoder → (μ, σ) → z

z → Decoder → x̂

Шығын бойынша Encoder мен Decoder жаңартылады.

Оқытудан кейін

z_жаңа ~ N(0, I)

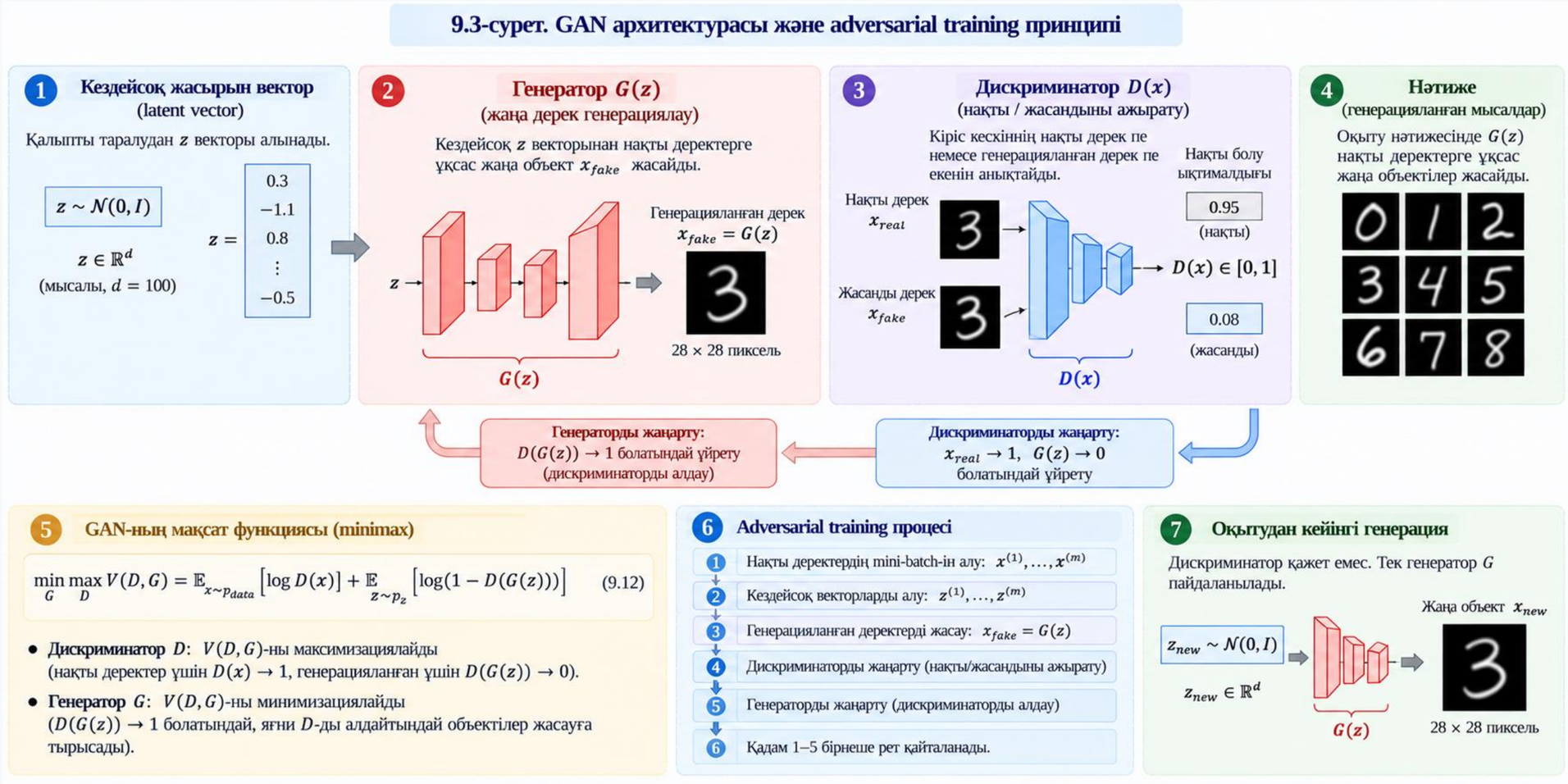
z_жаңа → Decoder → x_жаңа

Бастапқы x кірісі қажет емес.

Классикалық автоэнкодер: қайта құру.
VAE: ықтималдық жасырын құрылым арқылы генерацияны да ұйымдастыру.

Модельдің генерация сапасы оқыту деректеріне, архитектураға және мақсатқа тәуелді.

Generative Adversarial Network — GAN



Дереккөз: конспект, 5-бет, 9.3-сурет.

GAN: қарсылас оқыту мақсаты

Классикалық мақсат: min_G max_D V(D, G)

$$V(D, G) = E_x[\log D(x)] + E_z[\log(1 - D(G(z)))]$$

Discriminator D

Нақты дерек үшін D(x) → 1, генерацияланған үлгі үшін D(G(z)) → 0.

Generator G

Генерацияланған үлгіні D нақты деп қабылдайтындай етіп бейімделеді.

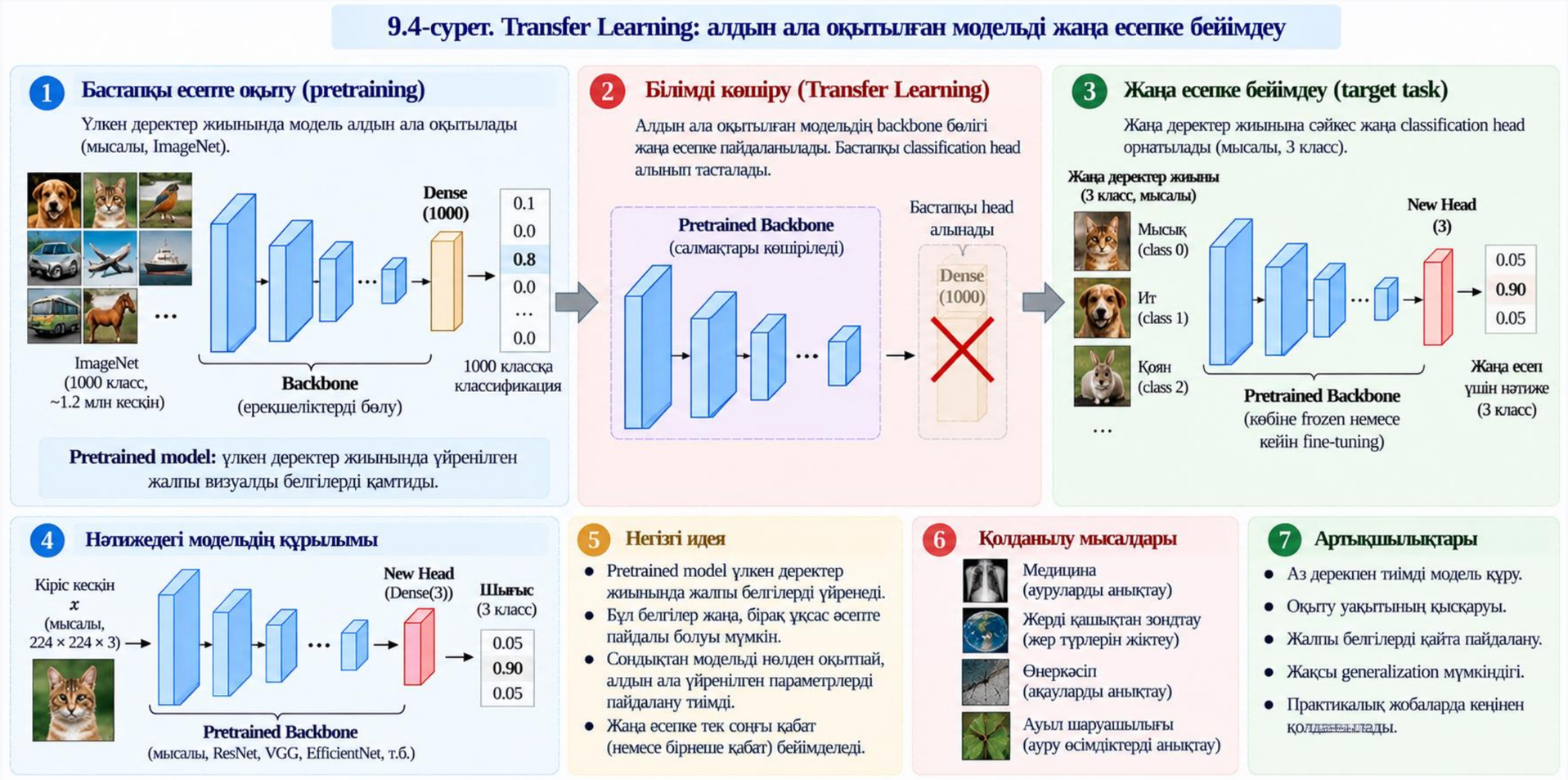
G және D кезектесіп оқытылады; практикада G үшін −E_z[log D(G(z))] жиі қолданылады.

VAE және GAN: таңдау және негізгі қиындықтар

Сипаттама	VAE	GAN
Құрылым	Encoder + Decoder	Generator + Discriminator
Оқыту негізі	Reconstruction + KL	Қарсылас мақсаттар
Классикалық Encoder	Бар	Міндетті емес

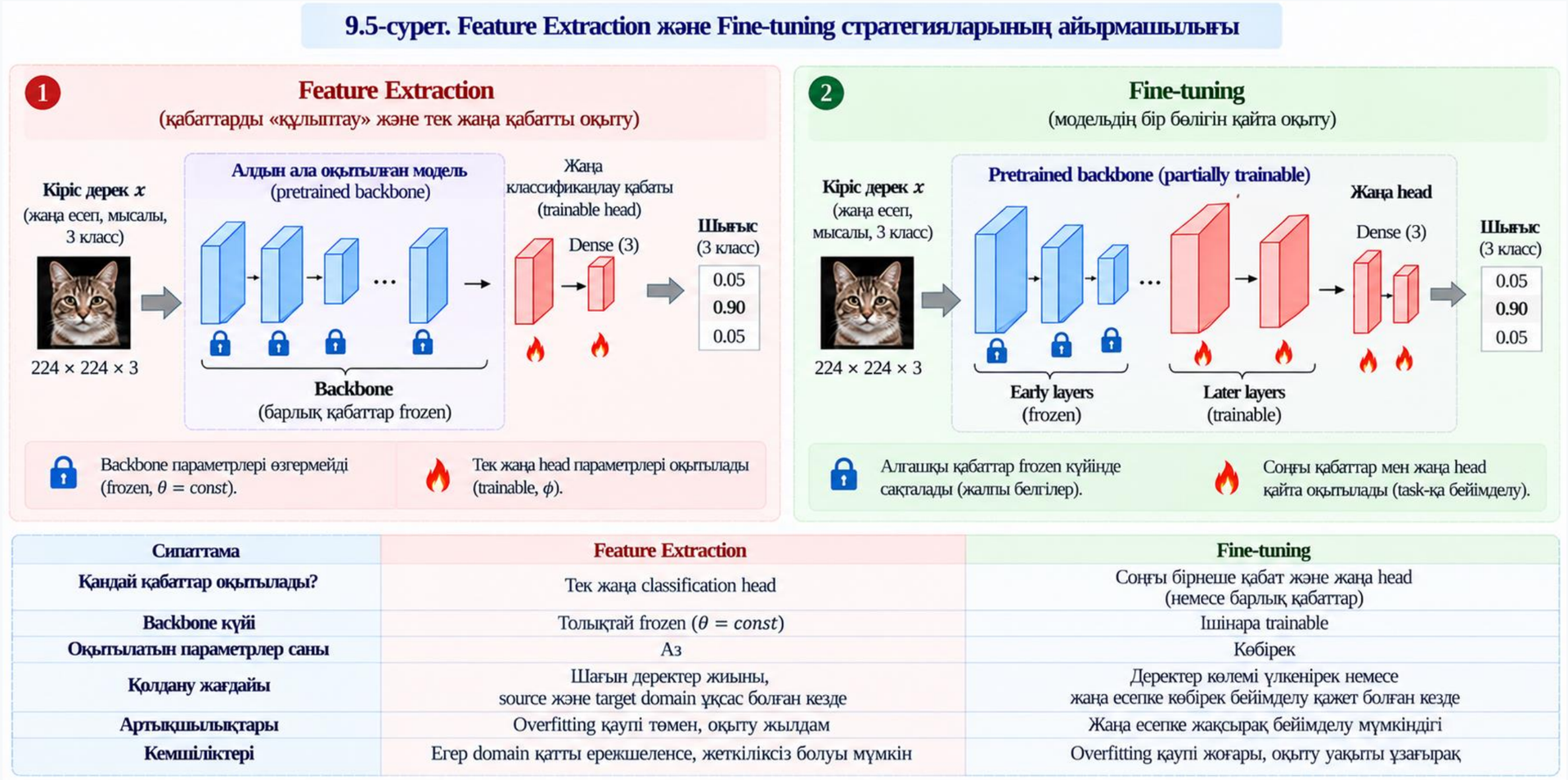
GAN қиындықтары: оқыту тұрақсыздығы және mode collapse.
Mode collapse: әртүрлі z мәндері өте ұқсас үлгілерге айналады.
Бағалау шынайылық пен әртүрлілікті қатар ескеруі тиіс.

Transfer Learning: білімді жаңа есепке көшіру



Дереккөз: конспект, 7-бет, 9.4-сурет.

Feature extraction және fine-tuning



Сипаттама	Feature Extraction	Fine-tuning
Қандай қабаттар оқытылады?	Тек жаңа classification head	Соңғы бірнеше қабат және жаңа head (немесе барлық қабаттар)
Backbone күйі	Толықтай frozen ($\theta = const$)	Ішінара trainable
Оқытылатын параметрлер саны	Аз	Көбірек
Қолдану жағдайы	Шағын деректер жиыны, source және target domain ұқсас болған кезде	Деректер көлемі үлкенірек немесе жаңа есепке көбірек бейімделу қажет болған кезде
Артықшылықтары	Overfitting қаупі төмен, оқыту жылдам	Жаңа есепке жақсырақ бейімделу мүмкіндігі
Кемшіліктері	Егер domain қатты ерекшеленсе, жеткіліксіз болуы мүмкін	Overfitting қаупі жоғары, оқыту уақыты ұзағырақ

Дереккөз: конспект, 8-бет, 9.5-сурет.

Қандай параметрлер оқытылады?

h = fθ(x) — backbone; ŷ = gφ(h) — жаңа head.

Feature extraction

θ = const: backbone бекітілген.

Тек φ параметрлері оқытылады:
φ ← φ − η∇φL.

Оқытылатын параметрлер саны аз.

Fine-tuning

Backbone-ның бір бөлігі не барлығы және жаңа head оқытылады.

θ_late ← θ_late − η_ft ∇θ_late L.

Көбіне кіші learning rate таңдалады.

Frozen — параметрлері жаңартылмайды; trainable — шығын бойынша бейімделеді.

Практикалық мысал: 1000 кластан 3 класқа

Backbone → Dense(1000) ⇒ Backbone → Dense(3)

1. Head ауыстыру	Жаңа {A, B, C} кластарына арналған шығыс қабатын орнату.
2. Бастапқы оқыту	Backbone-ды бекітіп, жаңа head параметрлерін оқыту.
3. Қажет болса	Соңғы қабаттарды ашып, төмен қадаммен fine-tuning жасау.

Transfer Learning стратегиясын таңдау

Жағдай	Бастапқы стратегия
Target дерегі шағын, domain ұқсас	Frozen backbone + жаңа head
Дерек көбірек, бейімдеу қажет	Соңғы қабаттарды fine-tuning
Domain едәуір өзгеше	Көбірек қабатты бейімдеуді тексеру

Negative transfer — бастапқы білім жаңа есептің нәтижесін нашарлатуы.

Соңғы шешім — validation нәтижелері бойынша; test модель таңдау үшін қолданылмайды.

Қолданылуы және нәтижені бағалау

Генеративті модельдер

Кескін, мәтін және дыбыс үлгілері.
Синтетикалық деректер.
Қалпына келтіру, image-to-image өңдеу.

Шынайылық, әртүрлілік және нақты деректерге сәйкестік тексеріледі.

Transfer Learning

Аз дерекпен жаңа модельді бейімдеу.
Оқыту уақыты мен ресурстарын үнемдеу мүмкіндігі.

Кіріс өлшемі мен алдын ала өңдеу pretrained модель талаптарына сәйкес болуы тиіс.

Жақсы көрінетін бірнеше үлгі немесе source Accuracy жалғыз өзі сапаны дәлелдемейді.

Ортақ негіз: үйренілген көріністі пайдалану

Генеративті оқыту

Жасырын құрылымды модельдеу.

z → жаңа объект

VAE және GAN осы идеяны әртүрлі оқыту мақсаттарымен іске асырады.

Білімді көшіру

Бұрын үйренілген көріністі жаңа міндетке бейімдеу.

x → pretrained representation → жаңа есеп

Frozen немесе fine-tuning режимі таңдалады.

Қазіргі генеративті модельдер VAE және GAN-мен шектелмейді: диффузиялық және Transformer негізіндегі модельдер де қолданылады.

Бақылау сұрақтары

01 p(y|x), p(x) және p(x|y) мақсаттары қалай ажыратылады?

02 VAE Encoder-і нені есептейді және қайта параметрлеу не үшін қажет?

03 Reconstruction және KL мүшелері VAE-де қандай міндет атқарады?

04 GAN-дағы G мен D қалай оқытылады, mode collapse нені білдіреді?

05 Feature extraction және fine-tuning кезінде қандай параметрлер өзгереді?

06 Domain сәйкестігі, деректер көлемі және negative transfer стратегия таңдауға қалай әсер етеді?

Дереккөз және иллюстрациялар

Негізгі дереккөз

«9-дәріс. Генеративті нейрондық желілер және Transfer Learning». Пайдаланушы ұсынған дәріс конспектісі, 10 бет.

Конспектідегі бес түпнұсқа схема сақталды

9.1-сурет: модель мақсаттары, 2-бет → 2-слайд.
9.2-сурет: VAE, 3-бет → 4-слайд.
9.3-сурет: GAN, 5-бет → 7-слайд.
9.4-сурет: Transfer Learning, 7-бет → 10-слайд.
9.5-сурет: екі бейімдеу стратегиясы, 8-бет → 11-слайд.

Формулалар конспект негізінде жүйеленді. Latent кластерлер схемалық түрде берілген; модель таңдау мен сапаны бағалау шарттары қысқа түсіндірмелермен нақтыланды.